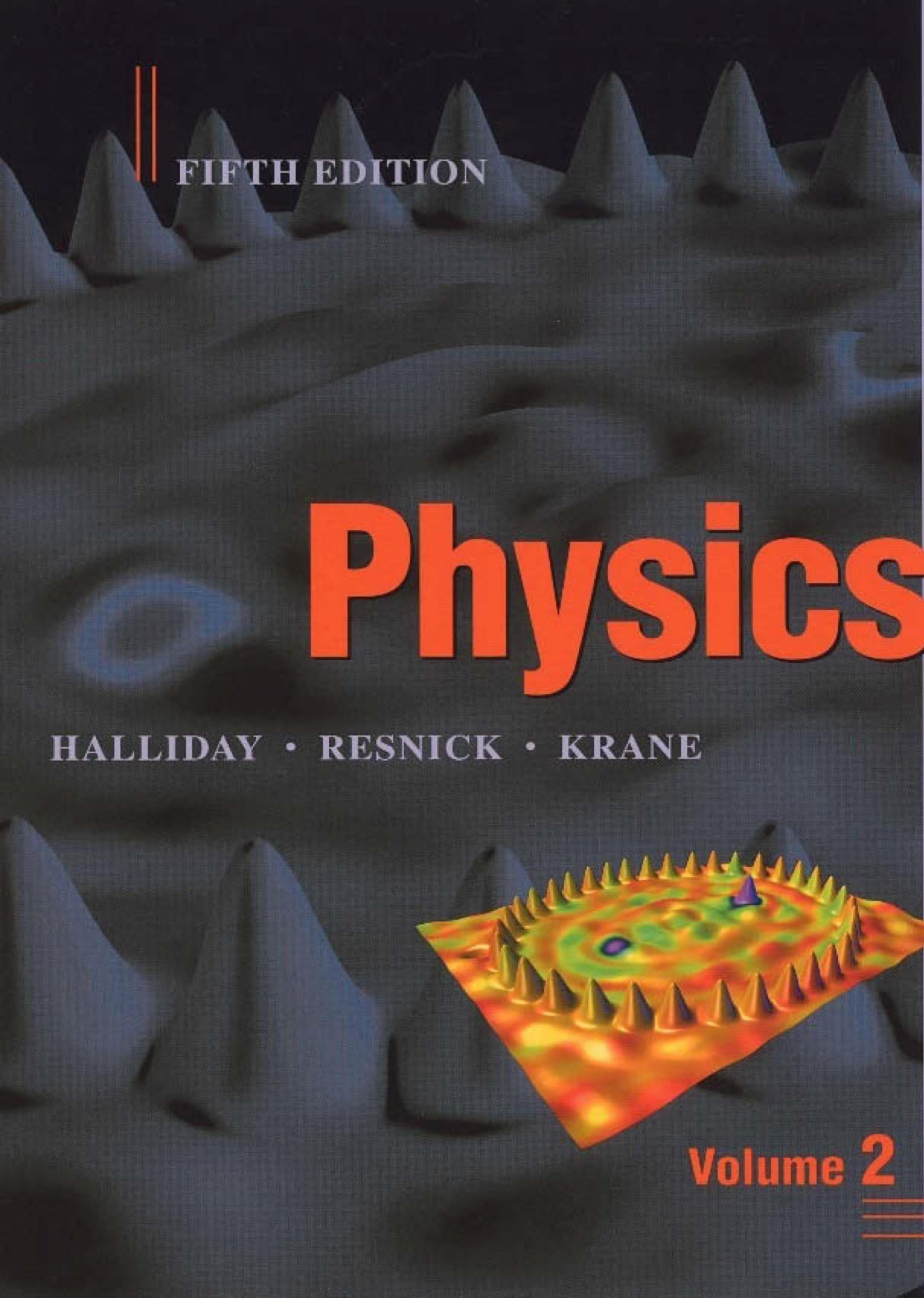


FIFTH EDITION

Physics

HALLIDAY · RESNICK · KRANE

Volume 2

SUPPLEMENTS

Instructor's Supplements

Instructor's Solutions Manual by PAUL STANLEY, California Lutheran University. This manual provides worked-out solutions for all of the end-of-chapter problems.

Instructor's Manual by J. RICHARD CHRISTMAN, U.S. Coast Guard Academy. This manual includes suggested syllabi, lecture notes, a list of the problems that appear in the Student Solutions Manual, a complete list of answers to the problems, a comparison of the problems with the Fourth Edition, and a list of computer projects.

Test Bank by J. RICHARD CHRISTMAN, U.S. Coast Guard Academy. This manual includes more than 2200 multiple-choice questions. These items are also available in the Computerized Test Bank (see below).

Instructor's Resource CD. This CD contains:
- All of the *Instructor's Solutions Manual* in both LaTex and pdf files
- Computerized Test Bank, in both IBM and Macintosh versions, with full editing features to help the instructor customize tests.
- All text illustrations, suitable for both classroom projection and printing.

Wiley Physics Simulations. This CD contains 50 interactive simulations covering all major topic areas in the introductory physics course. They are programmed in Java and can be used as lecture demonstrations or as on-line student assignments.

Wiley eGrade. eGrade is a powerful on-line homework management system that allows instructors to assign and grade homework using the web.

Student's Supplements

Study Guide by J. RICHARD CHRISTMAN, U.S. Coast Guard Academy. This student study guide provides an array of study aids and problem-solving help.

Student Solutions Manual by PAUL STANLEY, California Lutheran University. This manual provides students with complete worked-out solutions to 25 percent of the problems found at the end of each chapter of the text.

VOLUME TWO

PHYSICS

VOLUME TWO

PHYSICS

Fifth Edition

David Halliday
Professor of Physics
University of Pittsburgh

Robert Resnick
Professor of Physics
Rensselaer Polytechnic Institute

Kenneth S. Krane
Professor of Physics
Oregon State University

With the assistance of
Paul Stanley
California Lutheran University

JOHN WILEY & SONS, INC.

New York | **Chichester** | **Weinheim** | **Brisbane** | **Singapore** | **Toronto**

ACQUISITIONS EDITOR	Stuart Johnson
SENIOR PRODUCTION EDITOR	Elizabeth Swain
SENIOR MARKETING MANAGER	Bob Smith
SENIOR DESIGNER	Karin Gerdes Kincheloe
ILLUSTRATION EDITOR	Anna Melhorn
PHOTO EDITORS	Sara Wight and Hilary Newman
TEXT DESIGNER	Lee Goldstein
COVER DESIGNER	David Levy

Cover photo courtesy IBM Research, Almaden Research Center.

Line art created by Imagineering Art.

This book was set in 10/12 Times Roman by Progressive Information Technologies and was printed and bound by Courier Westford. The cover was printed by Lehigh.

This book is printed on acid-free paper. ∞

Copyright © 1960, 1962, 1966, 1978, 1992, 2002 John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

To order books or for customer service, call 1(800)-CALL-WILEY (225-5945).

Library of Congress Cataloging in Publication Data:

Halliday, David, 1916–
Physics. –– 5th ed. / David Halliday, Robert Resnick, Kenneth S. Krane.
p. cm
Resnick's name appears first on v. 1 of the 5th ed.; Halliday's name appears first on v. 2 of the 5th ed.
Includes bibliographical references and indexes.
ISBN 0-471-32057-9 (v. 1 : acid-free paper) – – ISBN 978-0-471-40194-0 (v. 2 : acid-free paper)
1. Physics. I. Resnick, Robert, 1923– II. Krane, Kenneth S. III. Title.
QC21.3 .H35 2002
530 –– dc21

2001017605

Printed in the United States of America

10 9 8 7 6 5

PREFACE TO VOLUME 2

This is the fifth edition of the textbook first published in 1960 as *Physics for Students of Science and Engineering* by David Halliday and Robert Resnick. For four decades this book has provided the standard for the calculus-based introductory survey course and has been known for the clarity and completeness of its presentation. In the present edition we have striven to increase accessibility without sacrificing the level or the rigor of its content. The text has been substantially rewritten to make the material flow more smoothly and to ease the student's entry into new subjects. We have attempted to provide more practical examples and to proceed from the particular to the general when new topics are introduced.

This edition features many changes in the pedagogy as well as in the ordering of material in the chapters. Those who are familiar with the fourth edition of this text will find the same topics but in a slightly revised order. In making these revisions, we have sought the advice of users of past editions and have taken into consideration the results of physics education research. Among the changes we have made in this edition are the following:

1. Owing to a rearrangement that resulted in the elimination of two chapters from Volume 1, the chapters in Volume 2 have been renumbered beginning with 25 (which corresponds to Chapter 27 of the fourth edition).

2. Students often have difficulties with integrating over continuous charge distributions in the calculation of electric fields, a procedure that is both conceptually abstract and computationally challenging. In order to deal earlier with the conceptual difficulties, we introduce the procedure in connection with electric *forces* rather than electric *fields;* for example, in Chapter 25 we discuss the calculation of the force exerted by a line of charge on a point charge. Students generally have a greater physical intuition for forces than fields, and in this way we can establish the mathematical procedure in a more physical context. Later we will repeat the calculations for fields and potentials. For similar reasons, we introduce the shell theorems in Chapter 25 in the context of forces, which parallels their introduction in Chapter 14 of Volume 1 in the discussion of the gravitational force.

3. The discussion of Rutherford scattering has been moved from the chapter on Gauss' law in the previous edition to the discussion of electric fields in Chapter 26.

4. In Chapter 27 (Gauss' Law), we have expanded the discussion of the relationship between electric flux and field lines, and we now discuss the conventional applications of Gauss' law to continuous charge distributions before its applications to conductors.

5. Chapter 29 (The Electrical Properties of Materials) is a new chapter that incorporates material on conductors and dielectrics that appeared in the previous edition in the two chapters on capacitors and current. We believe that this material stands on its own, and by introducing it in this way we can more easily contrast the behavior of conductors and insulators in electric fields.

6. Physics education research consistently shows that students have significant difficulties in understanding the behavior of simple DC circuits. We have therefore expanded our presentation of this topic while simultaneously decreasing the coverage of multiloop circuits and measuring instruments.

7. We now begin our introduction to the sources of the magnetic field (Chapter 33) with a presentation of the field due to a single moving charge, and then move to the field due to a current element. This enables a better correspondence with the way magnetic fields are introduced in the previous chapter (treating the force on a single moving charge first and then the force on a current element). We also now provide a direct calculation of the axial field of a solenoid using the Biot–Savart law before repeating the calculation using Ampère's law.

8. The introduction of the magnetic dipole moment has been delayed until Chapter 35 (Magnetic Properties of Materials). This has been done in part to avoid overloading students with new material in the first chapter on magnetic fields as well as to provide a more coherent approach by introducing the magnetic dipole in the context in which it will be most directly applied. We have shortened somewhat the discussion of atomic and nuclear magnetism here, preferring to delay a more detailed discussion until a later chapter following the introduction of some of the rudiments of atomic structure along with electron spin.

9. We have reconfigured Chapters 40, 41, 42, and 43 of the previous edition into Chapters 38 and 39 of the present edition. Chapter 38 now treats Maxwell's equations and their applications to electromagnetic waves, material that was included in Chapters 40 and 41 of the fourth edition. Chapter 39 introduces properties of light waves, including reflection and refraction, and thus incorporates material that previously appeared in Chapters 41, 42, and 43. Image formation by plane mirrors now appears in the following chapter (40), where it fits more naturally with the discussion of image formation by mirrors and lenses.

10. In the fourth edition, topics from modern physics were "sprinkled" throughout the text, generally in sections labeled as "optional." In this edition we continue to use examples from modern physics where appropriate throughout the text, but the separate sections on modern physics have been consolidated into Chapters 45–52, which treat topics from quantum physics and its applications to atoms, solids, and nuclei. We strongly believe that relativity and quantum physics are essential parts of an introductory survey course at this level, but that justice to these subjects is better accomplished by a coherent, unified presentation rather than a collection of isolated expositions. As was the case in the fourth edition, we continue to place the chapter on special relativity among the classical mechanics chapters in Volume 1, which reflects our strong belief that special relativity belongs squarely among the kinematics and mechanics chapters dealing with classical physics. Chapters 45–48, which treat quantum physics and its applications to atoms, have been substantially rewritten from the fourth edition. Chapter 45 introduces the usual early experiments suggesting the particle-like properties of electromagnetic radiation (thermal radiation, the photoelectric effect, Compton scattering). However, unambiguous evidence for the particle-wave duality of light comes only from modern delayed choice experiments, which we now also treat in Chapter 45. The rudiments of the Schrödinger theory are now treated in Chapter 46, with detailed applications to potential wells and to the hydrogen atom in Chapter 47. Chapter 48, which treats atomic structure, is similar to Chapter 52 of the fourth edition with the addition of new material on atomic magnetism.

The end-of-chapter material in this edition differs significantly from that of the previous edition. The previous problem sets (which were all keyed to chapter sections) have been carefully edited and placed into two groups: exercises and problems. Exercises, which are keyed to text sections, generally represent direct applications of the material in the associated section. Their purpose is usually to help students become familiar with the concepts, important formulas, units and dimensions, and so forth. Problems, which are not keyed to text sections, often require use of concepts from different sections or even from previous chapters. Some problems call for the student to estimate or independently to locate the data needed to solve the problem. In editing and grouping the exercises and problems, we have also eliminated some problems from the previous edition. A problem supplement will incorporate most of the missing problems as well as a selection of new exercises and problems. As before, answers to odd-numbered exercises and problems are given in the text and those to the even-numbered exercises and problems can be found in the Instructor's Manual that is available to instructors.

Multiple-choice questions and computer problems have also been added to the end-of-chapter material. The multiple-choice questions are generally conceptual in nature and often call for unusual insights into the material. Answers to the multiple-choice questions can be found in the instructor's manual. The computer problems may require familiarity with spread-sheet techniques or with symbolic manipulation routines such as Maple or Mathematica.

The development of the end-of-chapter material has been undertaken with the substantial assistance of Paul Stanley of California Lutheran University. We have been fortunate to have had the benefits of his insights and creativity in this project.

We have striven to develop a textbook that offers as complete and rigorous a survey of introductory physics as is possible at this level. It is, however, important to assert that *few (if any) instructors will want to follow the entire text from start to finish,* especially in a one-year course. There are many alternate pathways through this text. The instructor who wishes to treat fewer topics in greater depth (often called the "less is more" approach) will be able to select from among these pathways. Some sections or subsections are explicitly labeled as "optional," indicating that they can be skipped without loss of continuity. Depending on the course design, other sections or even entire chapters can be skipped or treated lightly. The Instructor's Manual, available as a companion volume, offers suggestions for abbreviating the coverage. Even so, the complete presentation remains in the text where the curious student can seek out the omitted topics and be rewarded with a broader view of the subject. We hope that the text can thus be regarded as a sort of "road map" through physics; many roads, scenic or direct, can be taken, and all roads need not be utilized on the first journey. The eager traveler may be encouraged to return to the map to explore areas missed on previous journeys.

The text is available in two volumes. Volume 1 covers kinematics, mechanics, and thermodynamics; the present volume covers electromagnetism, optics, and quantum physics and its applications. Supplements available include:

Instructor's Solutions Manual
Instructor's Manual
Instructor's Resource CD
Test Bank
Student Solutions Manual
Student Study Guide
Physics Simulations
eGrade Homework Management System

In preparing this edition, we have benefited from the advice of a dedicated team of reviewers who have, individually or collectively, carefully offered comments and criticisms on nearly every page of the text:

Richard Bukrey, Loyola University
Duane Carmony, Purdue University
J. Richard Christman, U. S. Coast Guard Academy
Paul Dixon, California State University-San Bernadino
John Federici, New Jersey Institute of Technology
David Gavenda, University of Texas-Austin
Stuart Gazes, University of Chicago
James Gerhart, University of Washington
John Gruber, San Jose State University
Martin Hackworth, Idaho State University
Jonathan Hall, Pennsylvania State University, Behrend
Oshri Karmon, Diablo Valley College
Jim Napolitano, Rensselaer Polytechnic Institute
Donald Naugle, Texas A&M University
Douglas Osheroff, Stanford University
Harvey Picker, Trinity College
Anthony Pitucco, Pima Community College
Robert Scherrer, Ohio State University
John Toutonghi, Seattle University

We are deeply indebted to these individuals for their efforts and for the insights they have provided to the authors. We would also like to acknowledge the advice of the Physics Education Group at the University of Washington, especially Paula Heron and Lillian McDermott.

The staff at John Wiley & Sons has provided constant support for this project, for which we are exceptionally grateful. We would especially like to thank Stuart Johnson for his management of this project and his dedication to its completion. Essential contributions to the quality of this text have been made by production editor Elizabeth Swain, photo editor Hilary Newman, illustration editor Anna Melhorn, and designer Karin Kinchloe. Without the skill and efforts of these individuals this project would not have been possible.

Despite the best efforts of authors, reviewers, and editors, it in inevitable that errors may appear in the text, and we welcome communication from users with corrections or comments on the content or pedagogy. We read all of these communications and respond to as many as possible, but we regret not being able to respond to all of them. Nevertheless, we encourage readers' comments, which can be sent to www.wiley.com/college/halliday.

CONTENTS

CHAPTER 25

ELECTRIC CHARGE AND COULOMB'S LAW 567

25-1 Electromagnetism: A Preview 567
25-2 Electric Charge 568
25-3 Conductors and Insulators 571
25-4 Coulomb's Law 573
25-5 Continuous Charge Distributions 576
25-6 Conservation of Charge 580
Questions and Problems 581

CHAPTER 26

THE ELECTRIC FIELD 587

26-1 What Is a Field? 587
26-2 The Electric Field 588
26-3 The Electric Field of Point Charges 590
26-4 Electric Field of Continuous Charge Distributions 592
26-5 Electric Field Lines 595
26-6 A Point Charge in an Electric Field 597
26-7 A Dipole in an Electric Field 600
26-8 The Nuclear Model of the Atom (Optional) 602
Questions and Problems 603

CHAPTER 27

GAUSS' LAW 611

27-1 What is Gauss' Law All About? 611
27-2 The Flux of a Vector Field 612
27-3 The Flux of the Electric Field 613
27-4 Gauss' Law 616
27-5 Applications of Gauss' Law 617
27-6 Gauss' Law and Conductors 621
27-7 Experimental Tests of Gauss' Law and Coulomb's Law 624
Questions and Problems 626

CHAPTER 28

ELECTRIC POTENTIAL ENERGY AND POTENTIAL 635

28-1 Potential Energy 635
28-2 Electric Potential Energy 636
28-3 Electric Potential 639
28-4 Calculating the Potential from the Field 640
28-5 Potential Due to Point Charges 641
28-6 Electric Potential of Continuous Charge Distributions 644
28-7 Calculating the Field from the Potential 646
28-8 Equipotential Surfaces 648
28-9 The Potential of a Charged Conductor 649
28-10 The Electrostatic Accelerator (Optional) 651
Questions and Problems 652

CHAPTER 29

THE ELECTRICAL PROPERTIES OF MATERIALS 661

29-1 Types of Materials 661
29-2 A Conductor in an Electric Field: Static Conditions 662
29-3 A Conductor in an Electric Field: Dynamic Conditions 663
29-4 Ohmic Materials 666
29-5 Ohm's Law: A Microscopic View 668
29-6 An Insulator in an Electric Field 670
Questions and Problems 672

CHAPTER 30

CAPACITANCE 679

30-1 Capacitors 679
30-2 Capacitance 679
30-3 Calculating the Capacitance 681
30-4 Capacitors in Series and Parallel 683
30-5 Energy Storage in an Electric Field 685
30-6 Capacitor with Dielectric 687
Questions and Problems 690

CHAPTER 31

DC CIRCUITS 701

31-1 Electric Current 701
31-2 Electromotive Force 703
31-3 Analysis of Circuits 704
31-4 Electric Fields in Circuits 709
31-5 Resistors in Series and Parallel 710
31-6 Energy Transfers in an Electric Circuit 713
31-7 *RC* Circuits 713
Questions and Problems 716

CHAPTER 32

THE MAGNETIC FIELD 725

32-1 Magnetic Interactions and Magnetic Poles 725
32-2 The Magnetic Force on a Moving Charge 727
32-3 Circulating Charges 731
32-4 The Hall Effect 734
32-5 The Magnetic Force on a Current-Carrying Wire 736
32-6 The Torque on a Current Loop 738
Questions and Problems 740

CHAPTER 33

THE MAGNETIC FIELD OF A CURRENT 749

33-1 The Magnetic Field due to a Moving Charge 749
33-2 The Magnetic Field of a Current 752
33-3 Two Parallel Currents 756
33-4 The Magnetic Field of a Solenoid 758
33-5 Ampère's Law 760
33-6 Electromagnetism and Frames of Reference (Optional) 764
Questions and Problems 765

CHAPTER 34

FARADAY'S LAW OF INDUCTION 775

34-1 Faraday's Experiments 775
34-2 Faraday's Law of Induction 776
34-3 Lenz' Law 777
34-4 Motional emf 780
34-5 Generators and Motors 782
34-6 Induced Electric Fields 783
34-7 Induction and Relative Motion (Optional) 786
Questions and Problems 789

CHAPTER 35

MAGNETIC PROPERTIES OF MATERIALS 801

35-1 The Magnetic Dipole 801
35-2 The Force on a Dipole in a Nonuniform Field 804
35-3 Atomic and Nuclear Magnetism 805
35-4 Magnetization 807
35-5 Magnetic Materials 808
35-6 The Magnetism of the Planets (Optional) 811
35-7 Gauss' Law for Magnetism 814
Questions and Problems 816

CHAPTER 36

INDUCTANCE 823

36-1 Inductance 823
36-2 Calculating the Inductance 824
36-3 *LR* Circuits 826
36-4 Energy Storage in a Magnetic Field 827
36-5 Electromagnetic Oscillations: Qualitative 830
36-6 Electromagnetic Oscillations: Quantitative 832
36-7 Damped and Forced Oscillations 833
Questions and Problems 836

CHAPTER 37

ALTERNATING CURRENT CIRCUITS 845

37-1 Alternating Currents 845
37-2 Three Separate Elements 846
37-3 The Single Loop *RLC* Circuit 848
37-4 Power in AC Circuits 851
37-5 The Transformer (Optional) 852
Questions and Problems 854

CHAPTER 38

MAXWELL'S EQUATIONS AND ELECTROMAGNETIC WAVES 861

38-1 The Basic Equations of Electromagnetism 861
38-2 Induced Magnetic Fields and the Displacement Current 862
38-3 Maxwell's Equations 864
38-4 Generating an Electromagnetic Wave 866
38-5 Traveling Waves and Maxwell's Equations 868
38-6 Energy Transport and the Poynting Vector 870
38-7 Radiation Pressure 872
Questions and Problems 874

CHAPTER 39

LIGHT WAVES 883

39-1 The Electromagnetic Spectrum 883
39-2 Visible Light 886
39-3 The Speed of Light 887
39-4 Reflection and Refraction of Light Waves 890
39-5 Total Internal Reflection 897
39-6 The Doppler Effect for Light 899
Questions and Problems 902

CHAPTER 40

MIRRORS AND LENSES 913

40-1 Image Formation by Mirrors and Lenses 913
40-2 Plane Mirrors 914
40-3 Spherical Mirrors 917
40-4 Spherical Refracting Surfaces 921
40-5 Thin Lenses 923
40-6 Optical Instruments 928
Questions and Problems 930

CHAPTER 41

INTERFERENCE 941

41-1 Two-Source Interference 941
41-2 Double-Slit Interference 942
41-3 Coherence 944
41-4 Intensity in Double-Slit Interference 946
41-5 Interference from Thin FIlms 948
41-6 Michelson's Interferometer 953
Questions and Problems 955

CHAPTER 42

DIFFRACTION 963

42-1 Diffraction and the Wave Theory of Light 963
42-2 Single-Slit Diffraction 965
42-3 Intensity in Single-Slit Diffraction 967
42-4 Diffraction at a Circular Aperture 970
42-5 Double-Slit Interference and Diffraction Combined 971
Questions and Problems 975

CHAPTER 43

GRATINGS AND SPECTRA 981

43-1 Multiple Slits 981
43-2 Diffraction Gratings 985
43-3 Dispersion and Resolving Power 986
43-4 X-ray Diffraction 988
43-5 Holography (Optional) 992
Questions and Problems 994

CHAPTER 44

POLARIZATION 999

44-1 Polarization of Electromagnetic Waves 999
44-2 Polarizing Sheets 1001
44-3 Polarization by Reflection 1003
44-4 Double Refraction 1004
44-5 Circular Polarization 1006
44-6 Polarization by Scattering 1008
Questions and Problems 1010

CHAPTER 45

THE NATURE OF LIGHT 1015

45-1 Introducing the Photon 1015
45-2 Thermal Radiation 1016
45-3 The Photoelectric Effect 1019
45-4 The Compton Effect 1021
45-5 The Photon Revealed 1023
45-6 Photons and Waves 1024
45-7 Slowing Down Atoms by Photon Bombardment 1026
Questions and Problems 1028

CHAPTER 46

THE NATURE OF MATTER 1035

46-1 Matter Waves 1035
46-2 Testing DeBroglie's Hypothesis 1036
46-3 Waves and Particles 1041
46-4 Heisenberg's Uncertainty Principle 1042
46-5 The Wave Function 1044
46-6 Schrödinger's Equation 1045
46-7 Barrier Tunneling 1046
Questions and Problems 1049

CHAPTER 47

ELECTRONS IN POTENTIAL WELLS 1055

47-1 Electrons, Free and Bound 1055
47-2 An Electron Trapped in a Potential Well 1055
47-3 An Electron Trapped in a Finite Well 1060
47-4 An Electron Trapped in an Atom 1062
47-5 The Ground State of the Hydrogen Atom 1065
47-6 Angular Momentum of Electrons in Atoms 1066
47-7 An Excited State of the Hydrogen Atom 1069
47-8 Counting the States of Hydrogen 1070
Questions and Problems 1072

CHAPTER 48

ATOMIC STRUCTURE 1079

48-1 The X-ray Spectrum of Atoms 1079
48-2 X Rays and the Numbering of the Elements 1081
48-3 Building Atoms 1082
48-4 The Periodic Table 1083
48-5 Atomic Magnetism 1086
48-6 The Stern–Gerlach Experiment 1087
48-7 Nuclear Magnetic Resonance 1089
48-8 Magnetism and Atomic Radiations (Optional) 1090
48-9 Lasers and Laser Light 1092
Questions and Problems 1096

CHAPTER 49

ELECTRICAL CONDUCTION IN SOLIDS 1103

49-1 Quantum Theory of Solids 1103
49-2 Conduction Electrons in a Metal 1104
49-3 Filling the Allowed States 1105
49-4 Electrical Conduction in Metals 1108
49-5 Bands and Gaps 1109
49-6 Conductors, Insulators, and Semiconductors 1111
49-7 Doped Semiconductors 1112
49-8 The *pn* Junction 1114
49-9 Optical Electronics 1117
49-10 The Transistor 1119
49-11 Superconductors 1120
Questions and Problems 1122

CHAPTER 50

NUCLEAR PHYSICS 1129

50-1 Discovering the Nucleus 1129
50-2 Some Nuclear Properties 1131
50-3 Radioactive Decay 1135
50-4 Alpha Decay 1136
50-5 Beta Decay 1138
50-6 Measuring Ionizing Radiation 1139
50-7 Natural Radioactivity 1140
50-8 Nuclear Reactions 1141
50-9 Nuclear Models (Optional) 1143
Questions and Problems 1145

CHAPTER 51

ENERGY FROM THE NUCLEUS 1153

51-1 The Atom and the Nucleus 1153
51-2 Nuclear Fission: The Basic Process 1154
51-3 Theory of Nuclear Fission 1155
51-4 Nuclear Reactors: The Basic Principles 1157
51-5 A Natural Reactor 1159
51-6 Thermonuclear Fusion: The Basic Process 1161
51-7 Thermonuclear Fusion in Stars 1162
51-8 Controlled Thermonuclear Fusion 1164
Questions and Problems 1167

CHAPTER 52

PARTICLE PHYSICS AND COSMOLOGY 1173

52-1 Particle Interactions 1173
52-2 Families of Particles 1176
52-3 Conservation Laws 1179
52-4 The Quark Model 1181
52-5 The Big Bang Cosmology 1186
52-6 Nucleosysthesis 1190
52-7 The Age of the Universe 1192
Questions and Problems 1194

APPENDICES

A. The International System of Units (SI) A-1
B. Fundamental Physcial Constants A-3
C. Astronomical Data A-4
D. Properties of the Elements A-6
E. Periodic Table of the Elements A-9
F. Elementary Particles A-10
G. Conversion Factors A-12
H. Vectors A-17
I. Mathematical Formulas A-20
J. Nobel Prizes in Physics A-22

ANSWERS TO ODD-NUMBERED PROBLEMS A-26

PHOTO CREDITS P-1

INDEX I-1

CHAPTER 25

ELECTRIC CHARGE AND COULOMB'S LAW

We begin here a detailed study of electromagnetism, *which will extend throughout most of the remainder of this text. Electromagnetic forces are responsible for the structure of atoms and for the binding of atoms in molecules and solids. Many properties of materials that we have studied so far are electromagnetic in their nature, such as the elasticity of solids and the surface tension of liquids. The spring force, friction, and the normal force all originate with the electromagnetic force between atoms.*

Among the examples of electromagnetism that we shall study are the force between electric charges, such as occurs between an electron and the nucleus in an atom; the motion of a charged body subject to an external electric force, such as an electron in an oscilloscope beam; the flow of electric charges through circuits and the behavior of circuit elements; the force between permanent magnets and the properties of magnetic materials; and electromagnetic radiation, which ultimately leads to the study of optics, *the nature and propagation of light.*

In this chapter we begin with a discussion of electric charge, some properties of charged bodies, and the fundamental electric force between two charged bodies.

25-1 ELECTROMAGNETISM: A PREVIEW

What do the following have in common?

1. You turn on the light switch in your room. The consumption of fuel at a power plant produces electromagnetic energy by causing a loop of electrically conducting wire to rotate in the vicinity of a magnet. Ultimately some of this energy is transferred to the electrons in the filament of your lightbulb, which can transform the electrical energy into visible light.

2. You enter a command on your computer keyboard. A stream of electrons is formed to transmit your instructions. There are many thousands of possible pathways for the electrons through the computer circuitry, but most are blocked by electronic gates. Electrons can move only through the gates that have been opened by your command so that the stream of electrons reaches its destination and your command is executed.

3. You push the channel select button on the remote control of your TV set. Electromagnetic waves travel from the remote control unit to a receptor on the set, which then tunes the set to accept another electromagnetic wave that originates from a satellite orbiting high above the Earth. The waves from the satellite provide instructions for your set to use electric and magnetic forces to focus and direct a beam of electrons that strikes the surface of the picture tube and produces a visible image.

The common factor in these diverse phenomena is that they all depend on forces that we describe as *electric* or *magnetic* to control and direct the flow of energy or particles. These forces form the basis of our study of *electromagnetism.* We will find in our study that all electromagnetic effects can be explained by a set of four basic equations, called *Maxwell's equations.* These equations represent individual laws of electromagnetism, just as we have previously discussed equations that represent Newtons' laws of mechanics or the laws of thermodynamics.

Our study will first consider electric phenomena and then magnetic phenomena. Later we will show that the two cannot be separated; certain electric phenomena produce magnetic effects, and certain magnetic phenomena produce electric effects. This leads us to unify electric and magnetic phenomena under the common name of electromagnetism. The development of the laws of electromagnetism and their unification was a great triumph of 19th-century physics. Their application has led directly to a great range of devices of practical use, such as motors, radios and televisions, radar, microwave ovens, and cellular phones.

The development of electromagnetic theory continued in the 20th century with three very significant advancements. In 1905, Albert Einstein showed that, to a moving observer, electric effects could appear as magnetic effects, and thus observers in relative motion could disagree in assigning their measurements to electric or magnetic causes. This conclusion formed the basis of the special theory of relativity, which ultimately was to revolutionize our concepts of space and time. The second development was the introduction of a quantum theory of electromagnetism, called *quantum electrodynamics,* which reached its fruition around 1949 and enabled properties of the atom to be calculated with incredible precision, currently about 11 significant figures. The third development of the 20th century was the unification of electromagnetism with another force, called the "weak" force, which is responsible for certain radioactive decay processes and other interactions between particles. Just as electric and magnetic effects were unified into the electromagnetic interaction, so electromagnetic and weak effects were shown in the 1960s to be unified under the *electroweak* interaction. For our study of electric and magnetic forces, however, the electroweak interaction does not yield anything new, and it is more convenient to consider the separate electromagnetic interaction.

Figure 25-1 is a time line of some of the major events in the development of our understanding of electromagnetism.

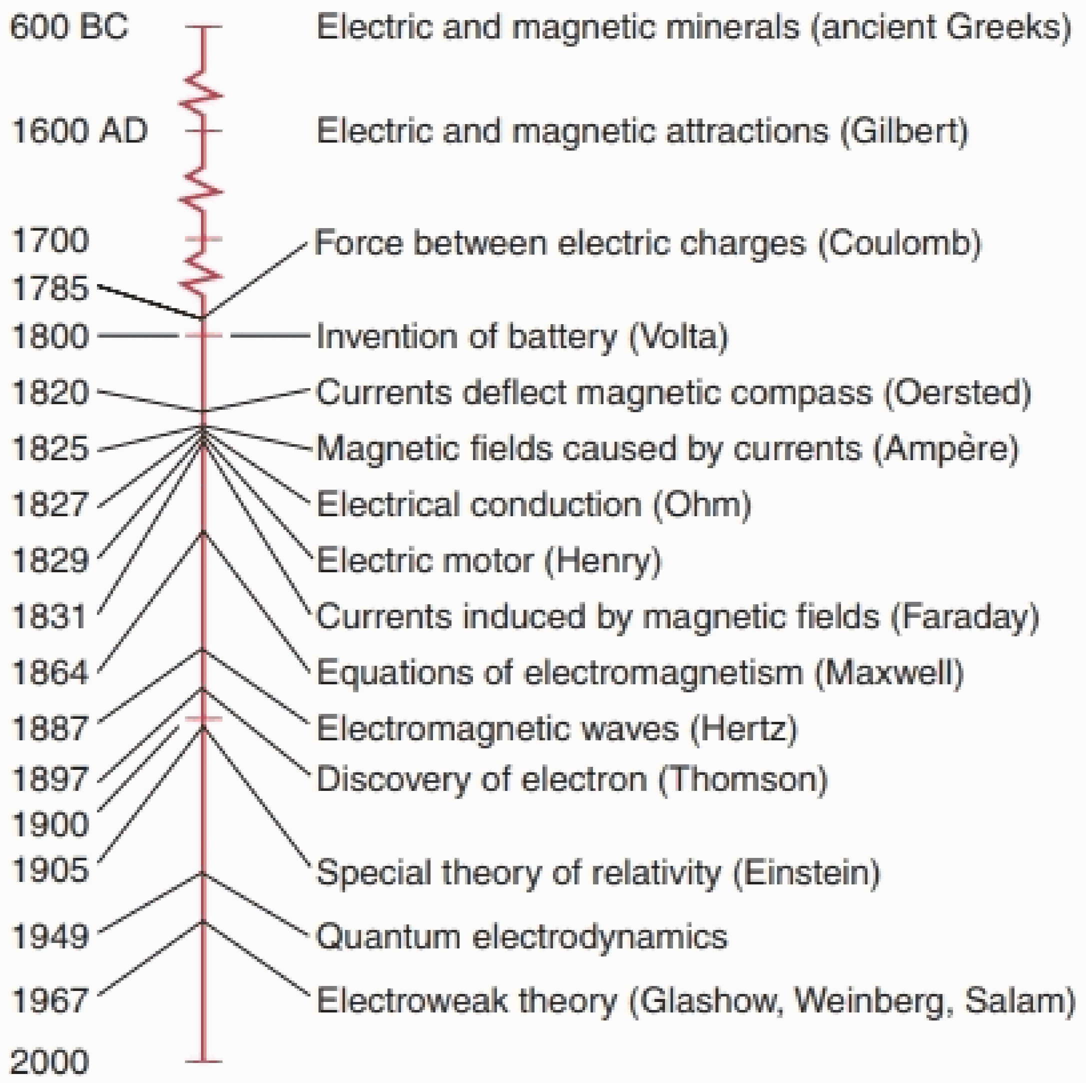

600 BC
Electric and magnetic minerals (ancient Greeks)
1600 AD
Electric and magnetic attractions (Gilbert)
1700
Force between electric charges (Coulomb)
1785
1800
Invention of battery (Volta)
1820
Currents deflect magnetic compass (Oersted)
1825
Magnetic fields caused by currents (Ampère)
1827
Electrical conduction (Ohm)
1829
Electric motor (Henry)
1831
Currents induced by magnetic fields (Faraday)
1864
Equations of electromagnetism (Maxwell)
1887
Electromagnetic waves (Hertz)
1897
Discovery of electron (Thomson)
1900
1905
Special theory of relativity (Einstein)
1949
Quantum electrodynamics
1967
Electroweak theory (Glashow, Weinberg, Salam)
2000

FIGURE 25-1. Time line of major developments in electromagnetism.

25-2 ELECTRIC CHARGE

After you pass a plastic comb through your hair a few times, you will find that the comb can exert a force on individual strands of your hair. You may also observe that, once the strands of hair are attracted to the comb and come into contact with it, they may no longer be attracted to it.

It seems reasonable to conclude that the attraction between the comb and the hair is a result of some physical entity being transferred from one to the other when they rub together, with the same physical entity being transferred back again to neutralize the attraction when they come into contact. This physical entity is called *electric charge,* and today we understand this transfer on the basis of electrons that can be removed from the atoms of one object and attached to the atoms of the other object.

The transfer of electric charge by means of friction is a commonly observed phenomenon. It was known to the ancient Greeks, who observed that pieces of amber rubbed with fur could attract bits of straw. When you walk across a carpet and are shocked by touching a metal door knob, or when a lightning flash stretches between a cloud and the ground, you are observing the effects of this transfer of charge.

When we "charge" an object (that is, when we transfer charge to it), we find that it can exert a force on another charged object. Early observations that this force can be either attractive or repulsive led to the conclusion that there are two kinds of electrical charge, which are called positive and negative.*

Although effects resulting from the transfer of charge can be powerful, it is remarkable that they originate from the transfer of only a tiny fraction of the electric charge that is contained in objects. Ordinary matter is made of electrically neutral atoms or molecules that contain equal amounts of positive charge (the nucleus) and negative charge (the electrons). When two objects rub together, relatively few electrons from the atoms of one object are

---

* The positive and negative labels for electric charge were chosen arbitrarily by Benjamin Franklin (1706–1790) who, among his other accomplishments, was a scientist of international reputation. In fact, his scientific reputation may have enabled his diplomatic triumphs in France during the American War of Independence.

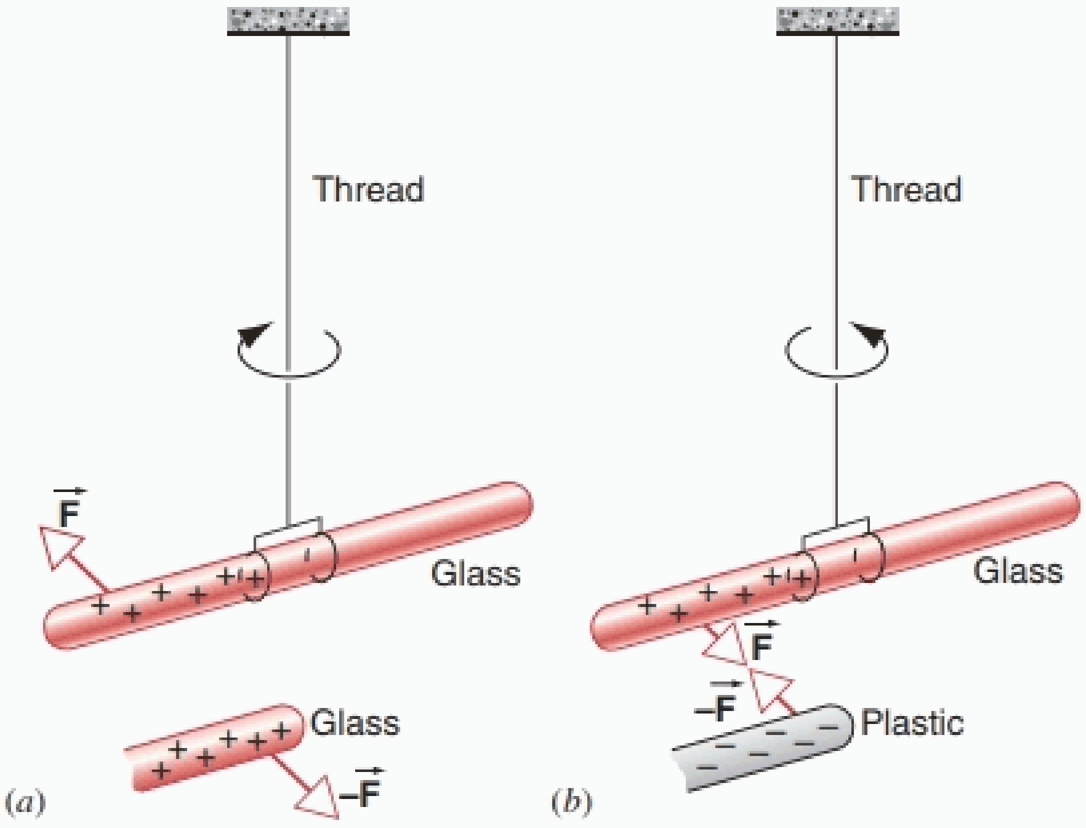


FIGURE 25-2. (*a*) Two similarly charged rods repel each other. (*b*) Two oppositely charged rods attract each other.

transferred to the other; most of the electrons remain undisturbed. It is this slight upset of the balance between the enormous but equal amounts of positive and negative charge in an object that is responsible for most commonly observed electrical effects.

When we rub a plastic rod with fur, electrons are transferred to the rod; because it has an excess of electrons (which carry a negative charge), the rod becomes negatively charged. The fur now has a deficiency of electrons and so it is positively charged. We can see the attraction of the rod for individual strands of the fur, which results from the charge on each. In a similar way, we can rub a glass rod with silk and observe that both become charged and can attract one another. In each case, we have transferred a relatively small number of electrons and upset the electrical neutrality of these objects.

Let us charge a glass rod by rubbing one end of it with silk and then suspend it from a thread, as in Fig. 25-2. If we place a similarly charged glass rod nearby, we find that the two rods repel one another, as in Fig. 25-2*a*. However, if we place a charged plastic rod (charged by rubbing with fur) nearby, the two rods attract one another, as in Fig. 25-2*b*.

We account for the existence of these two kinds of forces in terms of two kinds of charge. When plastic is rubbed with fur, electrons are transferred to the plastic and it becomes negatively charged. When glass is rubbed with silk, electrons are transferred to the silk, leaving the glass with a deficiency of electrons and therefore a net positive charge. The forces observed in Fig. 25-2 can be summarized by the following rule:

> *Charges of the same sign repel one another, and charges of the opposite sign attract one another.*

In Section 25-4, we put this rule into quantitative form, as Coulomb's law of force. We consider only charges that are either at rest with respect to each other or moving very slowly, a restriction that defines the subject of *electrostatics.*

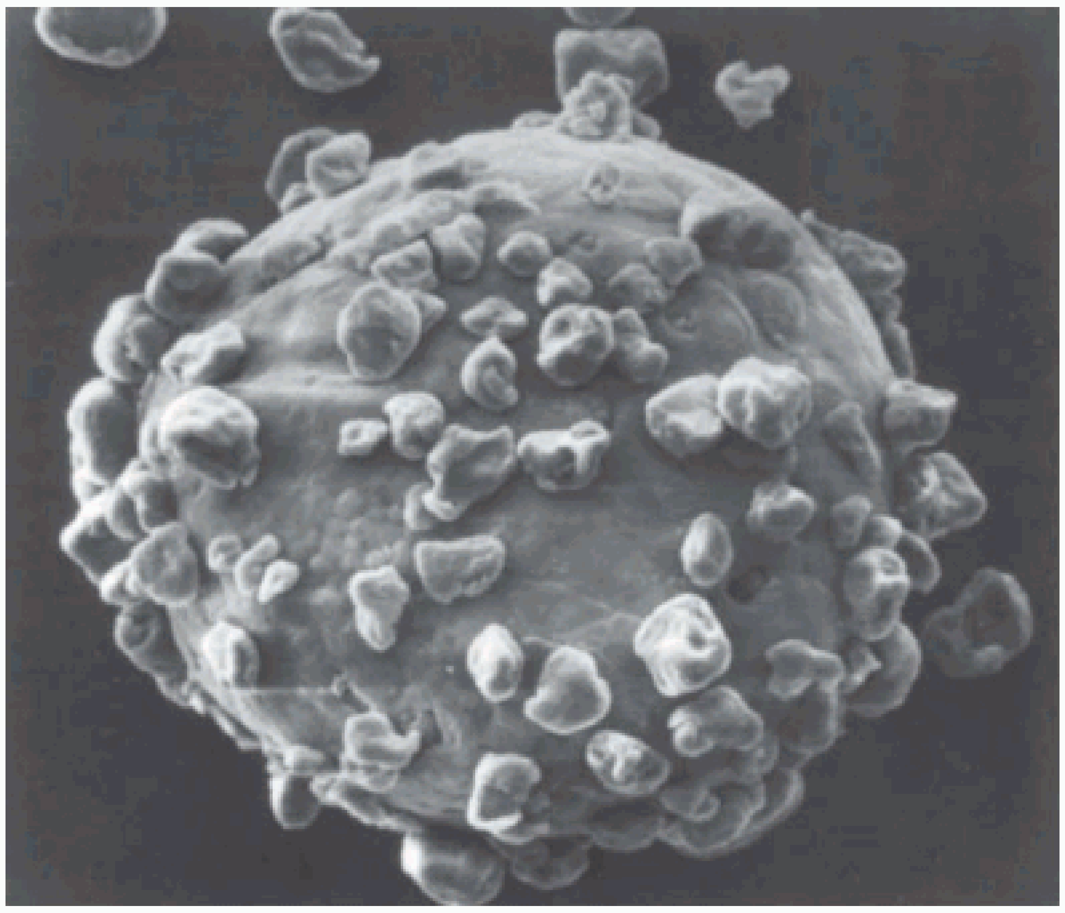


FIGURE 25-3. A carrier bead from a Xerox photocopier, covered with toner particles that stick to it by electrostatic attraction. The diameter of the bead is about 0.3 mm.

Electrical forces between charged bodies have many industrial applications, including electrostatic paint spraying and powder coating, fly-ash precipitation, nonimpact ink-jet printing, and photocopying. Figure 25-3 for example, shows a tiny carrier bead in a photocopying machine, covered with particles of black powder called *toner,* that stick to the carrier bead by electrostatic forces. These negatively charged toner particles are eventually attracted from their carrier beads to a positively charged latent image of the document to be copied, which is formed on a rotating drum. A charged sheet of paper then attracts the toner particles from the drum to itself, after which they are heat-fused in place to make the final copy.

The net electric charge of an object is usually represented by the symbol q. The charge is a scalar quantity. It can be positive or negative, depending on whether the object has a net positive or negative charge. Electric charge is measured in units of coulombs (C). The coulomb is a very large unit of charge; it takes about 6×10^{18} electrons to make one coulomb of charge.

The coulomb cannot be derived from previously defined units. Because electric charge is a new quantity, we are free to define its basic unit in any convenient way. One possible way would be in terms of the force exerted between two standard charges at a given separation, such as the quantity of charge that exerts a force of one newton on a similar charge a distance of one meter away. However, the force between static charges is difficult to measure, and so in practice it is more useful to define the coulomb in terms of the magnetic force between current-carrying wires (which is discussed in Chapter 33). This force can be measured more precisely than the electric force between static charges. It is therefore more convenient to define an SI base unit in terms of current (rate of flow of electric charge per unit time). The coulomb as a unit for electric charge is then a derived unit, obtained from the fundamental units of current and time (see Appendix A).

Electric Charge Is Quantized

When we transfer electric charge from one object to another, the transfer cannot be done in arbitrarily small units. That is, the flow of charge as a current is not a continuous flow, but is made up of discrete elements.* Experiments show that the electric charge always exists only in quantities that are integer multiples of a certain elementary quantity of charge e. That is,

$$q = ne \qquad n = 0, \pm 1, \pm 2, \pm 3, \ldots, \tag{25-1}$$

where (to four significant figures)

$$e = 1.602 \times 10^{-19}\ \mathrm{C}.$$

The *elementary charge e* is one of the fundamental constants of nature whose experimental value has been determined to an uncertainty of about 4 parts in 10^8.

The electron and the proton are examples of commonly occurring particles that each carry one fundamental unit of charge. The electron has a charge of $-e$ and the proton has a charge of $+e$. Some particles, such as the neutron, carry no net electric charge. Other elementary particles are known that carry charges that are small multiples of e, usually ± 1, ± 2, or ± 3. Each particle has a corresponding *antiparticle,* which has the same mass but the opposite electric charge; the antielectron, which is known as the *positron,* has a charge of $+e$. Antiparticles do not commonly exist in nature, but can be created in decays and reactions of nuclei and elementary particles.

Equation 25-1 tells us that it is possible to have a net charge on an object of $+10e$ or $-6e$ but never $3.57e$. When the values of a property are restricted to discrete multiples of a basic quantity, we say that the property is *quantized.*

Because the elementary charge is small, under ordinary circumstances we are not aware of the discrete nature of the flow of charge. For example, in a wire of an electronic circuit in which small currents of one milliampere are typical, 6×10^{15} electrons pass through any cross section of the wire every second!

Ordinary atoms are electrically neutral, which means that they contain equal quantities of positive and negative charge. The nucleus of the atom contains Z protons (where Z is called the *atomic number* of the atom) and thus a charge of $+Ze$. In a neutral atom, Z negatively charged electrons circulate about the nucleus. It is often possible to remove one or more electrons from an atom, creating an ion that has an excess positive charge of $+e$, $+2e$, ... For example, if we could remove all of the electrons from an atom of uranium ($Z = 92$), we would create a particle with a charge of $+92e$. Under certain circumstances, we can even attach an extra electron to a neutral atom, creating a negatively charged ion.

Although we believe electrons to be fundamental particles with no substructure, protons are not fundamental particles. They are made up of more elementary entities called *quarks.* The quarks are assigned fractional electric charges of $-\frac{1}{3}e$ and $+\frac{2}{3}e$. The proton is composed of three quarks, two with charges of $+\frac{2}{3}e$ and one with charge of $-\frac{1}{3}e$, which add up to a net charge of $+e$. Experimental evidence for the existence of quarks inside the proton is very strong (for example, high-energy electrons can be made to scatter from the fractionally charged quarks inside the proton), but, no matter how violently the protons are made to collide, no free quark has been released. As a result, no free particle with a fractional charge has ever been observed. This fact can be understood if the attractive force that one quark exerts on another *increases* with their separation. This is in contrast to the electromagnetic and gravitational forces, both of which *decrease* as the distance between a pair of interacting bodies increases.

SAMPLE PROBLEM 25-1. A penny, being electrically neutral, contains equal amounts of positive and negative charge. What is the magnitude of these equal charges?

Solution The charge q is given by NZe, in which N is the number of atoms in a penny and Ze is the magnitude of the positive and the negative charges carried by each atom.

The number N of atoms in a penny, assumed for simplicity to be made of copper, is $N_A m/M$, in which N_A is the Avogadro constant. The mass m of the coin is 3.11 g, and the mass M of 1 mol of copper (called its *molar mass*) is 63.5 g. We find

$$N = \frac{N_A m}{M} = \frac{(6.02 \times 10^{23}\ \text{atoms/mol})(3.11\ \text{g})}{63.5\ \text{g/mol}}$$
$$= 2.95 \times 10^{22}\ \text{atoms.}$$

Every neutral atom has a negative charge of magnitude Ze associated with its electrons and a positive charge of the same magnitude associated with its nucleus. Here e is the elementary charge, 1.60×10^{-19} C, and Z is the atomic number of the element in question. For copper, Z is 29. The magnitude of the total negative or positive charge in a penny is then

$$q = NZe = (2.95 \times 10^{22})(29)(1.60 \times 10^{-19}\ \mathrm{C})$$
$$= 1.37 \times 10^{5}\ \mathrm{C}.$$

This is an enormous charge. By comparison, the charge that you might get by rubbing a plastic rod is perhaps 10^{-9} C, smaller by a factor of about 10^{14}. For another comparison, it would take 1–2 days for a charge of 1.37×10^5 C to flow through the filament of a typical lightbulb. There is a lot of electric charge in ordinary matter.

---

* In Franklin's day, electric charge was thought to be a substance and to flow as a continuous fluid. Today we know that fluids are made up of individual atoms and molecules—matter is discrete. Similarly, the "electric fluid" is not continuous but discrete.

25-3 CONDUCTORS AND INSULATORS

Materials are commonly classified based on the ability of electrons to flow through them. In some materials, such as metals, electrons can move relatively freely. We call these materials *conductors.* Electrons deposited at one location in the material can easily move throughout the material. Other examples of conductors include tap water and the human body.

In other materials, electrons can hardly flow at all. Electrons deposited at one location will remain at that location. We call these materials *insulators.* Examples of insulators include glass, plastics, and many crystalline materials such as NaCl.

If you try to charge a copper rod by holding it in your hand and rubbing it with fur, you will not be successful. Electrons may be transferred from the rod to the fur as a result of friction, but additional electrons will easily flow from your body through your hand into the rod to replace those that were removed. As a result, no net charge builds up on the rod due to rubbing. We can consider the Earth to possess an infinite supply of electrons, some of which can flow into your body to replace those that have been lost to the rod. When we have a pathway through which electrons can flow between an object and the Earth, the object is said to be electrically *grounded.*

If instead we attach a plastic handle to the copper rod, we find that we can build up a charge by rubbing it. The insulating handle blocks the flow of electrons between the rod and your body.

Isolated atoms of a conducting material such as copper generally contain loosely bound electrons that can easily be detached, leaving a positively charged ion. When copper atoms come together to form solid copper, these loosely bound electrons do not remain attached to individual atoms but become free to wander through the material. These free electrons are called *conduction electrons;* in copper, a typical conductor, there are about 10^{23} conduction electrons per cm^3. The positively charged ions are not free to move and remain fixed within the solid lattice structure of the copper.

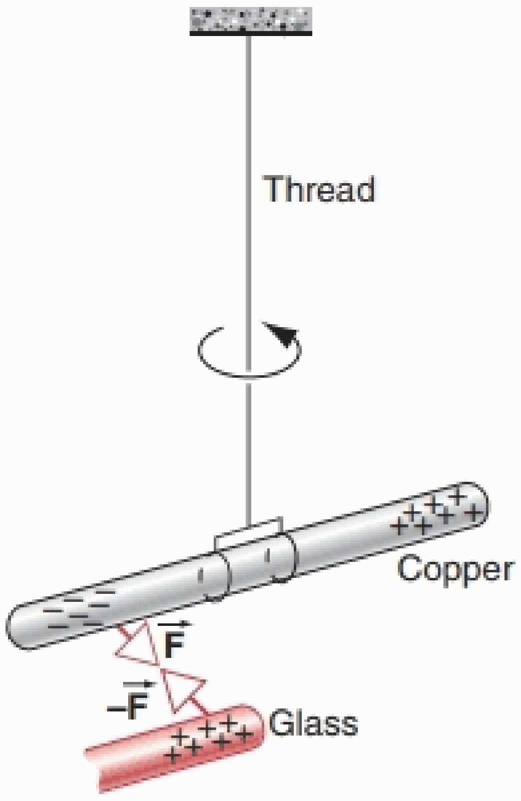


FIGURE 25-4. Either end of an isolated uncharged copper rod is attracted by a charged rod of either sign. In this case, conduction electrons in the copper rod are attracted to the near end of the copper rod, leaving the far end with a net positive charge.

The experiment of Fig. 25-4 demonstrates the mobility of charge in a conductor. An uncharged copper rod is suspended by an insulating thread. When a positively charged glass rod is brought near one end of the copper, the mobile conduction electrons in the copper are attracted by the positive charges on the glass. The flow of electrons to the near end of the copper leaves the far end with a deficiency of electrons and a net positive charge. The negatively charged end of the copper rod and the positively charged glass exert attractive forces on one another. Note that this situation is very different from that of Fig. 25-2; in Fig. 25-4, the glass attracts a copper rod that carries no net charge. (As we will discuss in the next section, the electrical force depends inversely on the separation between the charges; therefore the attractive force between the glass and the negative end of the copper rod is much stronger than the repulsive force between the glass and the positive end of the copper rod.)

If instead of a positively charged glass rod in Fig. 25-4 we used a negatively charged plastic rod, the effect would be the same: an attractive force between the plastic and the copper. In this case the negatively charged plastic would repel the conduction electrons of the copper, leaving the near end of the copper with a positive charge. There would be an attractive force between the negatively charged plastic and the positively charged end of the copper rod.

It is also possible to have an attractive force between a charged body and an uncharged insulator. Figure 25-5*a* shows a charged comb attracting uncharged bits of paper. The explanation for this attraction is different from that of the attraction between the glass rod and the copper rod. In this case, the paper is an insulator, and it is not possible for electrons to collect on one end of the paper (as was the case for the conductor in Fig. 25-4). Instead, electrons in individual molecules in the bits of paper are repelled by the negatively charged comb, and so the electrons are preferentially located on the side of each molecule that is away from the comb. In each molecule, the positive end (the end with a lack of electrons) is closest to the comb and feels a greater force of attraction to the comb. This is responsible for the net attractive force between the comb and the paper (Fig. 25-5*b*). The same attractive force would occur if the comb were positively charged.

The separation of positive and negative charge in an isolated object under the influence of a nearby charged object is known as *polarization.* Polarization can occur at the macroscopic level, such as in the copper rod of Fig. 25-4, or at the molecular level, as in Fig. 25-5.

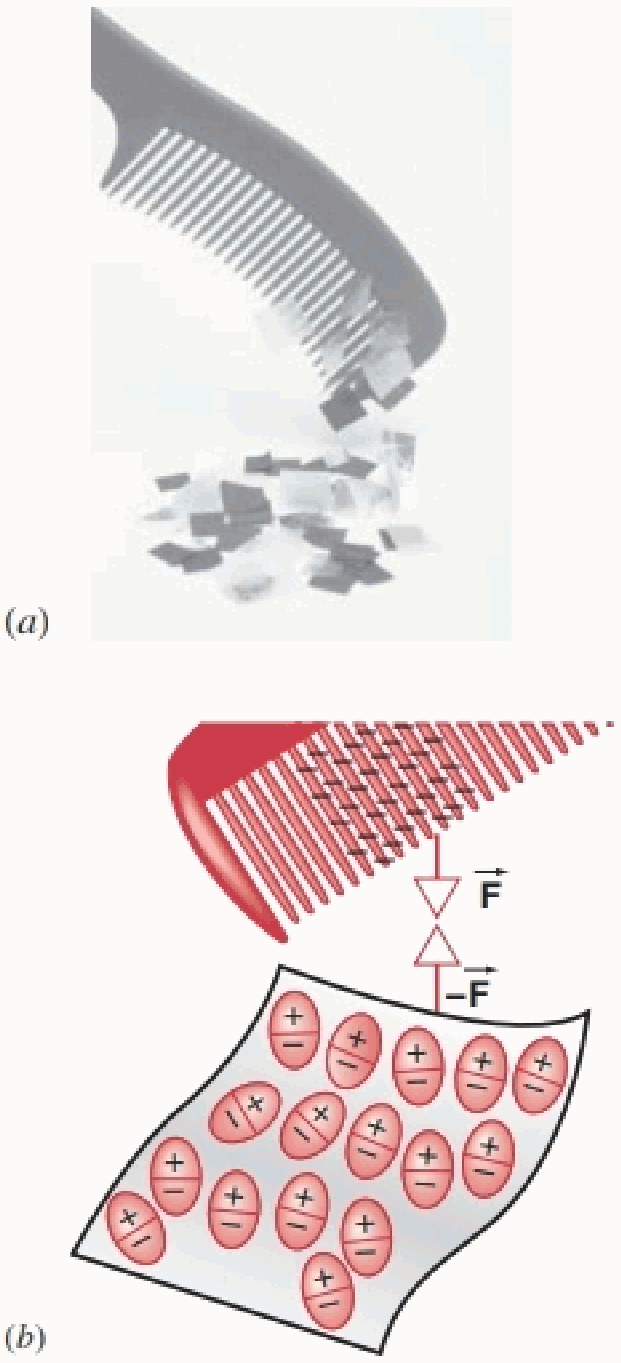


FIGURE 25-5. (*a*) A charged comb attracts uncharged bits of paper. (*b*) The negatively charged comb polarizes the charges in the molecules, resulting in an attractive force between the comb and the paper.

Charging by Contact and by Induction

Suppose now we touch a positively charged glass rod to an uncharged copper rod (as in Fig. 25-6). Electrons will flow from the copper to neutralize the positive charges on the glass. However, because electrons do not flow through the glass, they can neutralize only those positive charges at the point of contact with the copper. To remove additional electrons from the copper, we can "wipe" the glass rod along the copper rod, thereby transferring electrons to fresh, unneutralized areas of the glass that come into contact with the copper (Fig. 25-6*a*). If we then remove the glass rod, the copper remains with a deficiency of electrons and therefore a net positive charge. Electrons will flow through the copper so that the positive charges (the ion cores) are evenly distributed along the surface of the copper. This direct transfer of charge from one object to another is called *charging by contact.* Even though negative electrons were actually transferred, it is often convenient to regard the experiment shown in Fig. 25-6 as if positive charges were transferred from the glass rod to the copper.

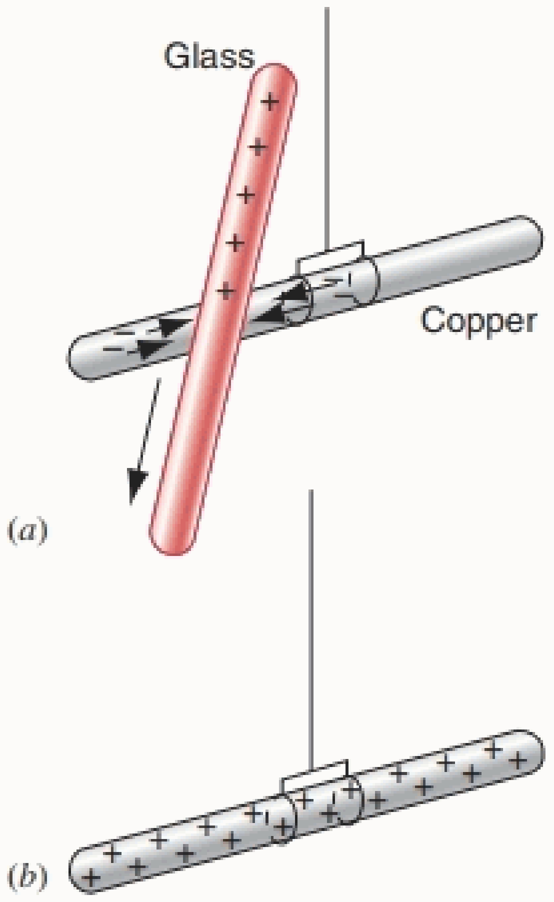


FIGURE 25-6. (*a*) Charging by contact. Electrons flow from the copper to neutralize positive charges at the point of contact with the glass. (*b*) The resulting charge on the copper when the glass is removed.

Let us return to the situation of Fig. 25-4. If we attach a wire between the positive end of the copper and ground (as in Fig. 25-7*a*), electrons will flow from ground to neutralize the positive charges in the copper. Leaving the glass rod in place, if we then remove the connection to ground, the copper rod retains a net negative charge. If we then remove the glass rod the negative charges will distribute themselves over the surface of the copper (Fig. 25-7*b*) to be as far apart from one another as possible. This method of charging an object is called *charging by induction.* Note that we have been able to use the positively charged glass rod to transfer either positive charge to the copper by contact or negative charge (from ground) by induction.

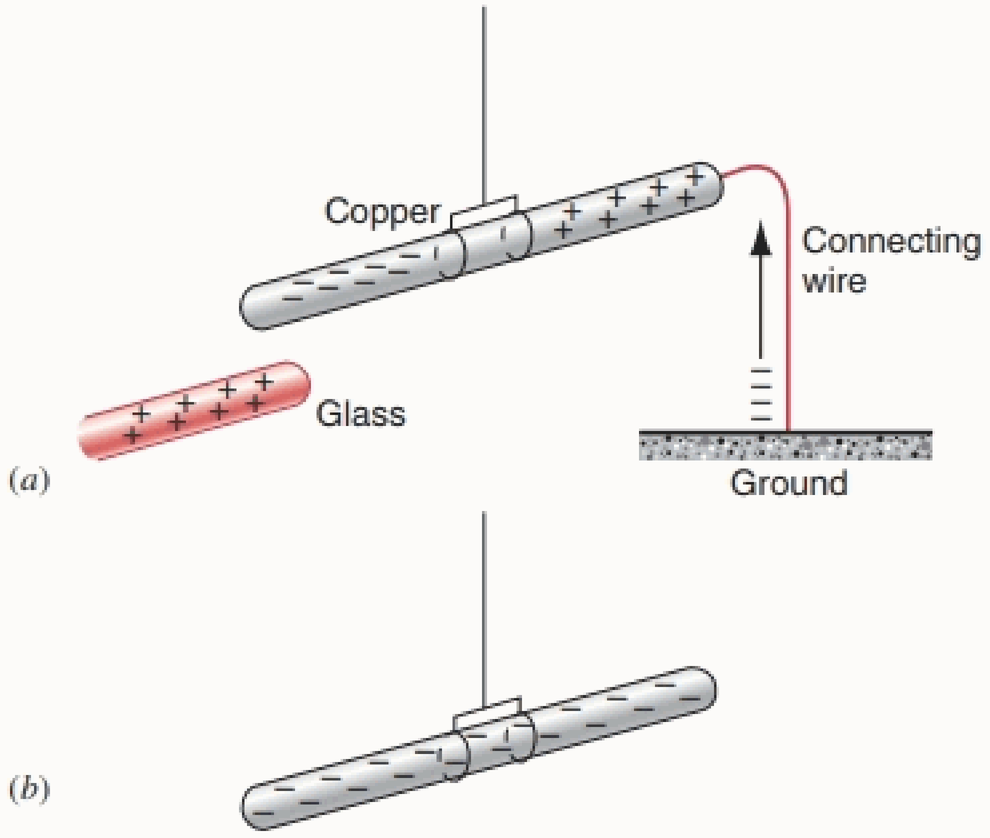


FIGURE 25-7. (*a*) Charging by induction. Electrons flow from ground to neutralize the positive charges on the far end of the copper rod. (*b*) The resulting charge on the copper when the glass is removed.

25-4 COULOMB'S LAW

So far in this chapter we have established that there are two kinds of electric charge and that charges exert forces on each other. It is now our goal to understand the nature of this force.

The first successful quantitative experiments to study the force between electric charges were done by Charles Augustin Coulomb (1736–1806), who measured electrical attractions and repulsions and deduced the law that governs them. In principle, Coulomb's apparatus is similar to Fig. 25-2, except that he used small charged spheres indicated by a and b in Fig. 25-8.

If a and b are charged, the electric force on a tends to twist the suspension fiber. Coulomb cancelled out this twisting effect by turning the suspension head through the angle θ needed to keep the two charges at a particular separation. The angle θ is then a relative measure of the electric force acting on charge a. The device of Fig. 25-8 is a *torsion balance;* a similar arrangement was used later by Cavendish to measure gravitational attractions (Section 14-3).

Experiments by Coulomb and his contemporaries showed that the electrical force exerted by one charged body on another depends directly on the product of the magnitudes of the two charges and inversely on the square of their separation. That is,

$$F \propto \frac{|q_1||q_2|}{r^2}.$$

Here F is the magnitude of the mutual force that acts on each of the two charges q_1 and q_2, and r is the distance between their centers. The force on each charge due to the other acts along the line connecting the charges. As required by Newton's third law, the force exerted by q_1 on q_2 is equal in magnitude but opposite in direction to the force exerted by q_2 on q_1, even though the magnitudes of the charges may be different.

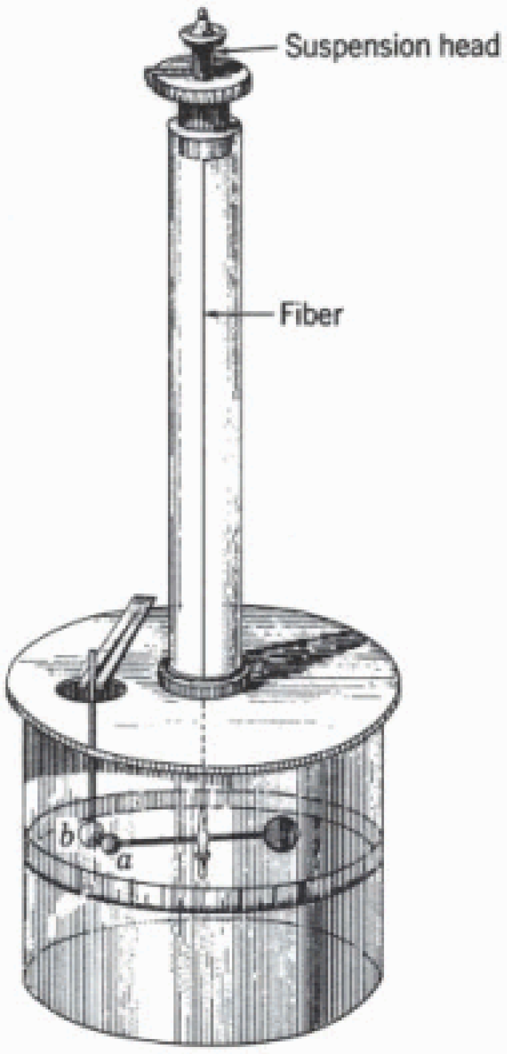


FIGURE 25-8. Coulomb's torsion balance, from his 1785 memoir to the Paris Academy of Sciences.

To turn the above proportionality into an equation, we introduce a constant of proportionality K, which we can call the Coulomb constant. We thus obtain, for the force between the charges,

$$F = K\frac{|q_1||q_2|}{r^2}. \tag{25-2}$$

Equation 25-2, which is called *Coulomb's law,* generally holds only for charged objects whose sizes are much smaller than the distance between them. We often say that it holds only for *point charges.*

Our belief in Coulomb's law does not rest quantitatively on Coulomb's experiments. Such measurements could not, for example, convince us that the exponent of r in Eq. 25-2 is exactly 2 and not, say 2.0001. In Section 27-7 we show that Coulomb's law can also be deduced from an indirect experiment, which shows that, if the exponent in Eq. 25-2 is not exactly 2, it differs from 2 by at most 1×10^{-16}.

Coulomb's law resembles Newton's inverse square law of gravitation, $F = Gm_1m_2/r^2$, which was already 100 years old at the time of Coulomb's experiments. Both are inverse square laws, and the charge q plays the same role in Coulomb's law that the mass m plays in Newton's law of gravitation. One difference between the two laws is that gravitational forces are always attractive, whereas electrostatic forces can be repulsive or attractive, depending on whether the two charges have the same or opposite signs.

In the SI system, the constant K is expressed in the following form:

$$K = \frac{1}{4\pi\epsilon_0}. \tag{25-3}$$

Although the choice of this form for the constant K appears to make Coulomb's law needlessly complex, it ultimately results in a simplification of formulas of electromagnetism that are used more often than Coulomb's law.

The constant ϵ_0, which is called the *electric constant* (also known as the *permittivity*), has a value that is determined by the value of the speed of light, as we discuss in Chapter 39. Its exact value is

$$\epsilon_0 = 8.85418781762 \times 10^{-12}\ \mathrm{C^2/N\cdot m^2}.$$

The Coulomb constant K has the corresponding value (to three significant figures)

$$K = \frac{1}{4\pi\epsilon_0} = 8.99 \times 10^9\ \mathrm{N\cdot m^2/C^2}.$$

With this choice of the constant K, Coulomb's law can be written

$$F = \frac{1}{4\pi\epsilon_0}\frac{|q_1||q_2|}{r^2}. \tag{25-4}$$

When K has the above value, expressing q in coulombs and r in meters gives the force in newtons.

The significance of Coulomb's law goes far beyond the description of the forces exerted by charged spheres on each other. This law, when incorporated into the structure of quantum physics, correctly describes (1) the electrical forces that bind the electrons of an atom to its nucleus, (2) the forces that bind atoms together to form molecules, and (3) the forces that bind atoms and molecules together to form solids or liquids. Thus most of the forces of our daily experience that are not gravitational in nature are electrical.

SAMPLE PROBLEM 25-2. In Sample Problem 25-1 we saw that a copper penny contains both positive and negative charges, each of a magnitude 1.37×10^5 C. Suppose that these charges could be concentrated into two separate bundles, held 100 m apart. What attractive force would act on each bundle?

Solution From Eq. 25-4 we have

$$F = \frac{1}{4\pi\epsilon_0}\frac{|q|^2}{r^2} = \frac{(8.99 \times 10^9\ \mathrm{N\cdot m^2/C^2})(1.37 \times 10^5\ \mathrm{C})^2}{(100\ \mathrm{m})^2}$$

$$= 1.69 \times 10^{16}\ \mathrm{N}.$$

This is about 2×10^{12} tons of force! Even if the charges were separated by one Earth diameter, the attractive force would still be about 120 tons. In all of this, we have sidestepped the problem of forming each of the separated charges into a "bundle" whose dimensions are small compared to their separation. Such bundles, if they could ever be formed, would be blasted apart by mutual Coulomb repulsion forces.

The lesson of this sample problem is that you cannot disturb the electrical neutrality of ordinary matter very much. If you try to pull out any sizable fraction of the charge contained in a body, a large Coulomb force appears automatically, tending to pull it back.

SAMPLE PROBLEM 25-3. The average distance r between the electron and the proton in the hydrogen atom is 5.3×10^{-11} m. (*a*) What is the magnitude of the average electrostatic force that acts between these two particles? (*b*) What is the magnitude of the average gravitational force that acts between these particles?

Solution (*a*) From Eq. 25-4 we have, for the electrostatic force,

$$F_e = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r^2} = \frac{(8.99 \times 10^9\ \mathrm{N\cdot m^2/C^2})(1.60 \times 10^{-19}\ \mathrm{C})^2}{(5.3 \times 10^{-11}\ \mathrm{m})^2}$$

$$= 8.2 \times 10^{-8}\ \mathrm{N}.$$

Although this force may seem small (it is about equal to the weight of a speck of dust), it produces an enormous acceleration of the electron within the atom, about $10^{23}\ \mathrm{m/s^2}$.

(*b*) For the gravitational force, we have

$$F_g = G\frac{m_e m_p}{r^2}$$

$$= \frac{(6.67 \times 10^{-11}\ \mathrm{N\cdot m^2/kg^2})(9.11 \times 10^{-31}\ \mathrm{kg})(1.67 \times 10^{-27}\ \mathrm{kg})}{(5.3 \times 10^{-11}\ \mathrm{m})^2}$$

$$= 3.6 \times 10^{-47}\ \mathrm{N}.$$

We see that the gravitational force is weaker than the electrostatic force by a factor of about 10^{39}. Although the gravitational force is weak, it is always attractive. Thus it can act to build up very large masses, as in the formation of stars and galaxies, so that large gravitational forces can develop. The electrostatic force, on the other hand, is repulsive for charges of the same sign, so that it is not possible to accumulate large concentrations of either positive or negative charge. We must always have the two types of charge together, so that they largely compensate for each other. The charges that we are accustomed to in our daily experiences are slight disturbances of this overriding balance.

SAMPLE PROBLEM 25-4. The nucleus of an iron atom has a radius of about 4×10^{-15} m and contains 26 protons. What repulsive electrostatic force acts between two protons in such a nucleus if they are separated by a distance of one radius?

Solution From Eq. 25-4 we have

$$F = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r^2} = \frac{(8.99 \times 10^9\ \mathrm{N\cdot m^2/C^2})(1.60 \times 10^{-19}\ \mathrm{C})^2}{(4 \times 10^{-15}\ \mathrm{m})^2}$$

$$= 14\ \mathrm{N}.$$

The large repulsive electrostatic force, more than 3 lb and acting on a single proton, must be balanced by the attractive nuclear force that binds the nucleus together. This force, whose range is so short that its effects cannot be felt very far outside the nucleus, is known as the "strong nuclear force" and is very well named.

Coulomb's Law: Vector Form

So far we have considered only the magnitude of the force exerted by one charge on another, determined according to Coulomb's law. Force, being a vector, has directional properties as well. In the case of Coulomb's law, the direction of the force is determined by the relative sign of the two electric charges.

As illustrated in Fig. 25-9, suppose we have two point charges q_1 and q_2 separated by a distance r_{12}. For the moment, we assume the two charges to have the same sign, so that they repel one another. Let us consider the force *on* particle 1 exerted *by* particle 2, which we write in our usual form as $\vec{\mathbf{F}}_{12}$. The position vector that locates particle 1 relative to particle 2 is $\vec{\mathbf{r}}_{12}$; that is, if we were to define the origin of our coordinate system at the location of particle 2, then $\vec{\mathbf{r}}_{12}$ would be the position vector of particle 1.

If the two charges have the same sign, then the force is repulsive and, as shown in Fig. 25-9*a*, $\vec{\mathbf{F}}_{12}$ must be parallel to $\vec{\mathbf{r}}_{12}$. If the charges have opposite signs, as in Fig. 25-9*b*, then the force $\vec{\mathbf{F}}_{12}$ is attractive and antiparallel to $\vec{\mathbf{r}}_{12}$. In either case, we can represent the force as

$$\vec{\mathbf{F}}_{12} = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r_{12}^2}\,\hat{\mathbf{r}}_{12}. \tag{25-5}$$

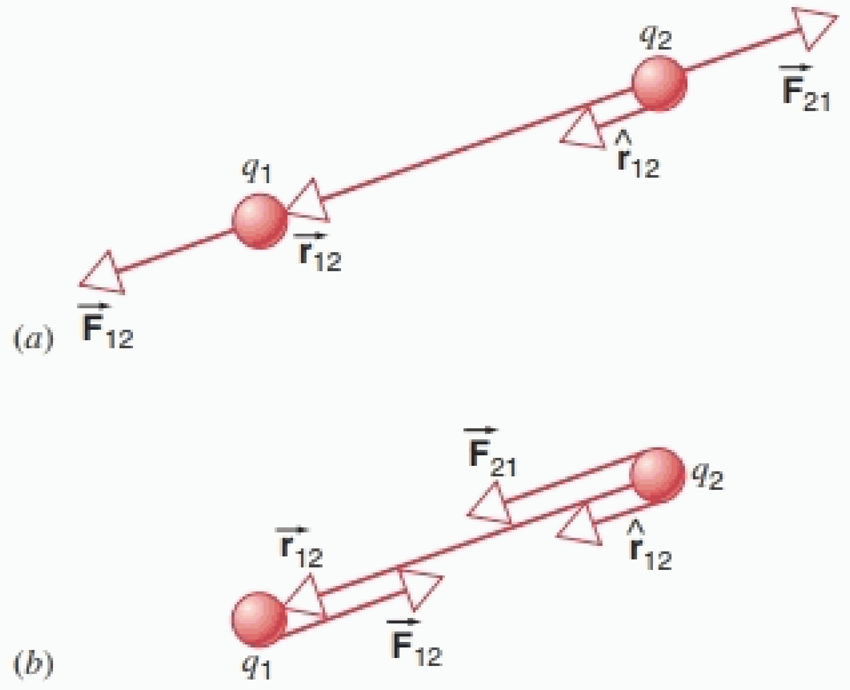


FIGURE 25-9. (*a*) Two point charges q_1 and q_2 of the same sign exert equal and opposite repulsive forces on one another. The vector $\vec{\mathbf{r}}_{12}$ locates q_1 relative to q_2, and the unit vector $\hat{\mathbf{r}}_{12}$ points in the direction of $\vec{\mathbf{r}}_{12}$. Note that $\vec{\mathbf{F}}_{12}$ is parallel to $\vec{\mathbf{r}}_{12}$. (*b*) The two charges now have opposite signs, and the force is attractive. Note that $\vec{\mathbf{F}}_{12}$ is antiparallel to $\vec{\mathbf{r}}_{12}$.

Here r_{12} represents the magnitude of the vector $\vec{\mathbf{r}}_{12}$, and $\hat{\mathbf{r}}_{12}$ indicates the *unit vector* in the direction of $\vec{\mathbf{r}}_{12}$. That is,

$$\hat{\mathbf{r}}_{12} = \frac{\vec{\mathbf{r}}_{12}}{r_{12}}. \tag{25-6}$$

We used a form similar to Eq. 25-5 to express the gravitational force (see Eqs. 14-2 and 14-3).

One other feature is apparent from Fig. 25-9. According to Newton's third law, the force exerted *on* particle 2 *by* particle 1, $\vec{\mathbf{F}}_{21}$, is opposite to $\vec{\mathbf{F}}_{12}$. This force can then be expressed in eactly the same form:

$$\vec{\mathbf{F}}_{21} = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r_{21}^2}\,\hat{\mathbf{r}}_{21}. \tag{25-7}$$

Here $\hat{\mathbf{r}}_{21}$ is a unit vector that points from particle 1 to particle 2; that is, it would be the unit vector in the direction of particle 2 if the origin of coordinates were at the location of particle 1.

The vector form of Coulomb's law is useful because it carries within it the directional information about $\vec{\mathbf{F}}$ and whether the force is attractive or repulsive. Using the vector form is of critical importance when we consider the forces acting on an assembly of more than two charges. In this case, Eq. 25-5 would hold for every pair of charges, and the total force on any one charge would be found by taking the *vector* sum of the forces due to each of the other charges. For example, the force on particle 1 in an assembly would be

$$\vec{\mathbf{F}}_1 = \vec{\mathbf{F}}_{12} + \vec{\mathbf{F}}_{13} + \vec{\mathbf{F}}_{14} + \cdots, \tag{25-8}$$

where $\vec{\mathbf{F}}_{12}$ is the force on particle 1 from particle 2, $\vec{\mathbf{F}}_{13}$ is the force on particle 1 from particle 3, and so on. Equation 25-8 is the mathematical representation of the *principle of superposition* applied to electric forces. It asserts that the force acting on one charge due to another is independent of whether or not other charges are present, and therefore we can calculate the force separately for each pair of charges and then take their vector sum to find the net force on any charge. For instance, the force $\vec{\mathbf{F}}_{13}$ that particle 3 exerts on particle 1 is completely unaffected by the presence of particle 2. The principle of superposition is not at all obvious and may fail in the case of very strong electric forces. Only through experiment can its applicability be verified. For all situations we meet in this text, however, the principle of superposition is valid.

SAMPLE PROBLEM 25-5. Figure 25-10 shows three charged particles, held in place by forces not shown. What electrostatic force, due to the other two charges, acts on q_1? Take $q_1 = -1.2\ \mu\text{C}$, $q_2 = +3.7\ \mu\text{C}$, $q_3 = -2.3\ \mu\text{C}$, $r_{12} = 15$ cm, $r_{13} = 10$ cm, and $\theta = 32°$.

Solution This problem calls for the use of the superposition principle. We start by computing the magnitudes of the forces that q_2 and q_3 exert on q_1. We substitute the magnitudes of the charges into Eq. 25-5:

$$\begin{aligned} F_{12} &= \frac{1}{4\pi\epsilon_0}\frac{|q_1||q_2|}{r_{12}^2} \\ &= \frac{(8.99\times10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)(1.2\times10^{-6}\ \text{C})(3.7\times10^{-6}\ \text{C})}{(0.15\ \text{m})^2} \\ &= 1.77\ \text{N}. \end{aligned}$$

The charges q_1 and q_2 have opposite signs so that the force exerted by q_2 on q_1 is attractive. Hence $\vec{\mathbf{F}}_{12}$ points to the right in Fig. 25-10.

We also have

$$\begin{aligned} F_{13} &= \frac{(8.99\times10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)(1.2\times10^{-6}\ \text{C})(2.3\times10^{-6}\ \text{C})}{(0.10\ \text{m})^2} \\ &= 2.48\ \text{N}. \end{aligned}$$

These two charges have the same (negative) sign so that the force exerted by q_3 on q_1 is repulsive. Thus $\vec{\mathbf{F}}_{13}$ points as shown in Fig. 25-10.

The components of the resultant force $\vec{\mathbf{F}}_1$ acting on q_1 are determined by the corresponding components of Eq. 25-8, or

$$\begin{aligned} F_{1x} &= F_{12x} + F_{13x} = F_{12} + F_{13}\sin\theta \\ &= 1.77\ \text{N} + (2.48\ \text{N})(\sin 32°) = 3.08\ \text{N} \end{aligned}$$

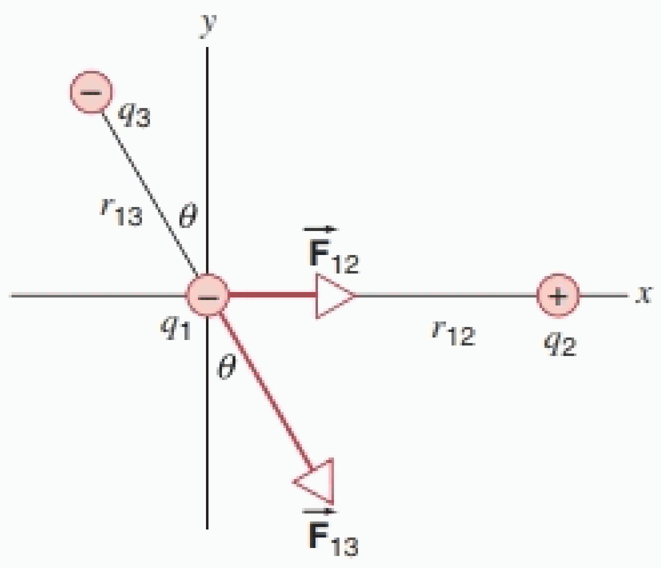


FIGURE 25-10. Sample Problem 25-5. The three charges exert three pairs of action–reaction forces on each other. Only the two forces acting on q_1 are shown here.

and

$$F_{1y} = F_{12y} + F_{13y} = 0 - F_{13}\cos\theta$$
$$= -(2.48\ \text{N})(\cos 32°) = -2.10\ \text{N}.$$

From these components, you can show that the magnitude of $\vec{\mathbf{F}}_1$ is 3.73 N and that this vector makes an angle of $-34°$ with the x axis.

25-5 CONTINUOUS CHARGE DISTRIBUTIONS

So far we have seen how to calculate the forces due to point charges. In many applications, however, electric forces are exerted by charged objects in the form of rods, plates, or solids. For simplicity we assume that the objects are insulators and that charge is spread throughout the surface or volume of the object, forming a *continuous charge distribution.*

Figure 25-2 showed forces exerted by one charged rod on another. Coulomb's law applies only to point charges, and we therefore cannot use Coulomb's law in its point charge form to calculate the force exerted by one charged rod on the other. It is possible to imagine the rods to be covered with point charges and to use Coulomb's law to calculate the force exerted by each point charge of one rod on each point charge of the other, but such an approach would be hopelessly complicated—if the rods carry the small charge of only 1 nC, it would be necessary to consider 10^{10} point charges on each rod!

Instead, we return to an idea from Franklin's time and regard the charge as a continuous property. The basic procedure is to divide the charge into infinitesimal elements and use the methods of calculus to find the total force due to all the elements.

If an object contains a net charge q, we imagine it to be divided into many small elements dq. Each element has a certain length, area, or volume, depending on whether we are considering charges that are respectively distributed in one, two, or three dimensions. We express dq in terms of the size of the element and the *charge density,* which describes how the charges are distributed over the length, area, or volume of the object. For most of the problems we consider in this text, the charges will be distributed uniformly over the object, which means that the charge density has the same value everywhere on the object.

In some situations, charges are distributed in one dimension, such as the thin charged rods of Fig. 25-2. In this case, we express dq in terms of the *linear charge density* (charge per unit length) λ, whose basic unit is C/m. A small element of the rod of length dx has a charge dq given by

$$dq = \lambda\, dx. \tag{25-9}$$

If the rod is uniformly charged, so that a total charge q is spread evenly over its length L, then $\lambda = q/L$. For example, a rod of length $L = 0.12$ m carrying a uniformly distributed charge of $q = 5.4 \times 10^{-6}$ C has a linear charge density of $\lambda = q/L = 4.5 \times 10^{-5}$ C/m. A short piece of the rod of length $dx = 1.0$ mm would then carry a charge $dq = \lambda\, dx = 4.5 \times 10^{-8}$ C.

In other situations, charge might be distributed over a two-dimensional area, such as the surface of the carrier bead in Fig. 25-3. In this case, dq is expressed in terms of the *surface charge density* (charge per unit area) σ, measured in the SI unit of C/m^2. A small element of area dA would then have a charge given by

$$dq = \sigma\, dA. \tag{25-10}$$

If a charge q is spread uniformly over a surface of area A, then $\sigma = q/A$.

The charge also might be spread throughout the volume of a three-dimensional object. In this case, we use the *volume charge density* (charge per unit volume) ρ, whose SI unit is C/m^3. The charge dq in an element of volume dV would then be

$$dq = \rho\, dV. \tag{25-11}$$

If the charge q is distributed uniformly throughout the volume V, then $\rho = q/V$.

To illustrate these concepts, we will calculate expressions for the force exerted by a continuous charge distribution on a point charge q_0. By extending these methods, it is possible to calculate the force exerted by one continuous charge distribution on another.

The procedure for finding the force exerted by a continuous charge distribution on a point charge is as follows:

1. Imagine the continuous charge distribution to be divided into a large number of small charge elements.

2. Pick an arbitrary charge element and express its charge dq in terms of Eq. 25-9, 25-10, or 25-11, depending respectively on whether the charge is distributed over a line, an area, or a volume.

3. Because dq is infinitesimally small, we can treat it as a point charge. Express the magnitude of the force element dF exerted by the charge dq on the charge q_0 in terms of Coulomb's law, Eq. 25-4:

$$dF = \frac{1}{4\pi\epsilon_0}\frac{|dq||q_0|}{r^2}, \tag{25-12}$$

where r is the distance between dq and q_0.

4. By taking into account the signs and locations of dq and q_0, determine the direction of the force element $d\vec{\mathbf{F}}$.

5. The total force is then found by adding all the infinitesimal force elements, which involves the integral

$$\vec{\mathbf{F}} = \int d\vec{\mathbf{F}}. \tag{25-13}$$

In performing this integral, we usually need to take into account that different charge elements dq may give force elements $d\vec{\mathbf{F}}$ in different directions. Equation 25-13 really means three different equations for the three components of $\vec{\mathbf{F}}$:

$$F_x = \int dF_x, \qquad F_y = \int dF_y, \qquad F_z = \int dF_z. \quad (25\text{-}14)$$

Occasionally we can use arguments based on symmetry to avoid calculating one or two of these integrals.

A Uniform Line of Charge

Figure 25-11 shows a thin rod of length L that lies along the z axis and carries a uniformly distributed positive charge q, so that its linear charge density is $\lambda = q/L$. We want to find the force exerted by the rod on a positive point charge q_0 located on the perpendicular bisector of the rod (the positive y axis) a distance y from its center.

The figure shows the results of performing steps 1, 2, and 3 of our procedure. We imagine the rod divided into small elements of length dz. An arbitrary element of charge $dq = \lambda\, dz$ is located a distance z from its center and exerts a force dF on q_0, where

$$dF = \frac{1}{4\pi\epsilon_0}\frac{q_0 dq}{r^2}.$$

The direction of the force $d\vec{\mathbf{F}}$ is shown in the figure. There is no component of $d\vec{\mathbf{F}}$ in the x direction (perpendicular to the page), so $F_x = 0$. We can also use a symmetry argument to show that $F_z = 0$. For every charge element dq located at position $+z$, there is another charge element dq located at $-z$. When we add the forces due to the charge elements at $+z$ and $-z$, we find that the z components have equal magnitudes but point in opposite directions, so their sum is zero. Because the charge q_0 is located in the median plane of the rod, this cancellation will occur for every such pair of charge elements along the entire length of the rod. We therefore conclude that $F_z = 0$.

Only F_y remains to be calculated. The element $dF_y = dF \cos\theta$ is shown in Fig. 25-11. With $dq = \lambda\, dz$, $r^2 = y^2 + z^2$, and $\cos\theta = y/r$, we have

$$dF_y = dF\cos\theta = \frac{1}{4\pi\epsilon_0}\frac{q_0\lambda\, dz}{(y^2+z^2)}\frac{y}{\sqrt{y^2+z^2}}$$

$$F_y = \int dF_y = \frac{1}{4\pi\epsilon_0} q_0\lambda y \int_{-L/2}^{L/2} \frac{dz}{(y^2+z^2)^{3/2}}.$$

Evaluating the integral (see Appendix I and note that y is a constant), we obtain

$$F_y = \frac{1}{4\pi\epsilon_0}\frac{q_0 q}{y\sqrt{y^2 + L^2/4}}. \quad (25\text{-}15)$$

This force is in the positive y direction when q_0 and q are positive. If the charge q_0 is moved to other locations in the xy plane, the expression for the force may change (see Exercise 14).

It is often instructive to evaluate expressions such as this one in various limiting cases. Let us consider the result when $y \gg L$, in which case the force becomes

$$F_y \approx \frac{1}{4\pi\epsilon_0}\frac{q_0 q}{y^2},$$

which is just the expression for the force of one point charge on another. When we are very far from the charged rod, or when the rod is very small, it looks like a point charge.

A Ring of Charge

Figure 25-12 shows a thin ring of radius R carrying a uniformly distributed positive charge q, so that its linear charge density is $\lambda = q/2\pi R$. We wish to find the force exerted by the ring on a positive point charge q_0 located on the axis of the ring (which we choose as the positive z axis) a distance z from the center of the ring. A small charge

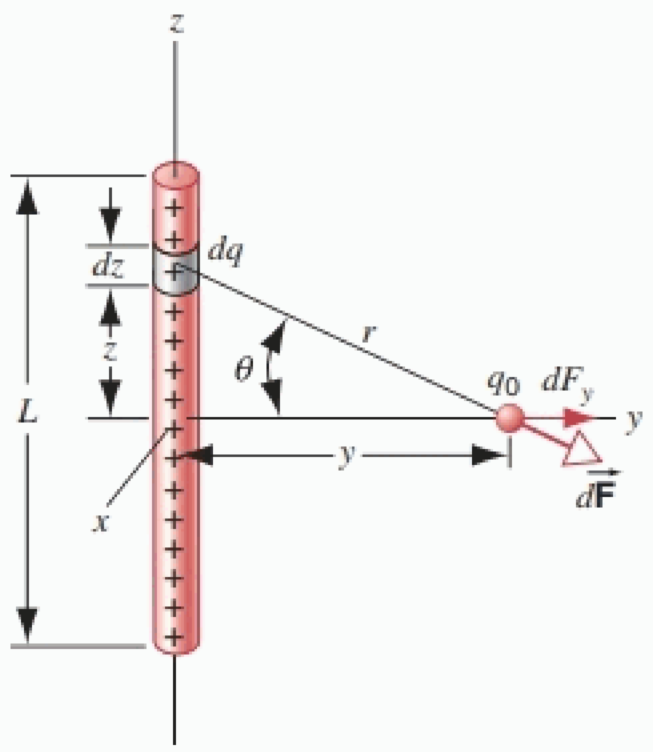


FIGURE 25-11. A uniformly charged rod. To find the force on the point charge q_0, we consider the rod to consist of many individual charge elements such as dq.

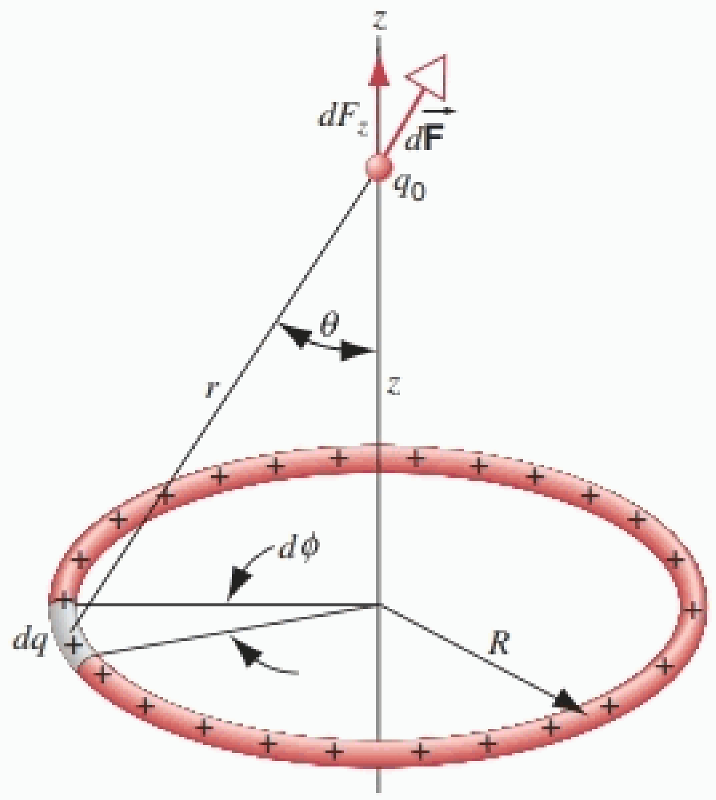


FIGURE 25-12. A uniformly charged ring. To find the force on a point charge q_0, we consider the ring to consist of many individual charge elements such as dq.

element of the ring has a length $R\,d\phi$ and thus carries a charge $dq = \lambda R\,d\phi$. The force dF exerted on q_0 by dq is

$$dF = \frac{1}{4\pi\epsilon_0}\frac{q_0\,dq}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{q_0\lambda R\,d\phi}{(z^2+R^2)}.$$

We can use a symmetry argument to establish that the only nonzero component of $\vec{\mathbf{F}}$ is its z component. For every element dq of the ring, there will be another element of equal charge dq at the opposite end of a diameter through the center of the ring; when the force elements on q_0 due to these two charge elements are added, all force components other than F_z will cancel. With $\cos\theta = z/r$, we find

$$\begin{aligned} F_z &= \int dF_z = \int dF\cos\theta \\ &= \int \frac{1}{4\pi\epsilon_0}\frac{q_0\lambda R\,d\phi}{(z^2+R^2)}\frac{z}{\sqrt{z^2+R^2}} \\ &= \frac{1}{4\pi\epsilon_0}\frac{q_0\lambda R z}{(z^2+R^2)^{3/2}}\int_0^{2\pi} d\phi. \end{aligned}$$

The integral evaluated around the ring gives 2π, and so the final result for the force is

$$F_z = \frac{1}{4\pi\epsilon_0}\frac{q_0 q z}{(z^2+R^2)^{3/2}}. \qquad (25\text{-}16)$$

Is this result also valid if q_0 is located on the negative z axis? (See Exercise 15.)

We can examine this result in the limiting case as $z \to \infty$. For $z \gg R$, we obtain

$$F_z \approx \frac{1}{4\pi\epsilon_0}\frac{q_0 q}{z^2},$$

which again gives the result for point charges. When we are very far from the ring, it appears to be a point charge.

Note also that $F_z = 0$ for $z = 0$. This is reasonable, because at the center of the ring the charge q_0 would be pushed equally in all directions by the charge elements that make up the ring.

A Disk of Charge

Figure 25-13 shows a circular disk of radius R that carries a positive charge q uniformly distributed over its surface, so that its surface charge density is $\sigma = q/\pi R^2$. A positive point charge q_0 is located on the axis of the disk (the positive z axis) a distance z from the center of the disk. To find the force exerted by the disk on the point charge, we can divide the disk into a series of concentric rings. The charge on the ring of radius w and thickness dw shown in Fig. 25-13 is $dq = \sigma\,dA = \sigma(2\pi w\,dw) = 2\pi\sigma w\,dw$. The force dF_z on q_0 exerted by this ring can be found using Eq. 25-16 with q replaced by dq and R replaced with w:

$$dF_z = \frac{1}{4\pi\epsilon_0}\frac{q_0(2\pi\sigma w\,dw)z}{(z^2+w^2)^{3/2}}.$$

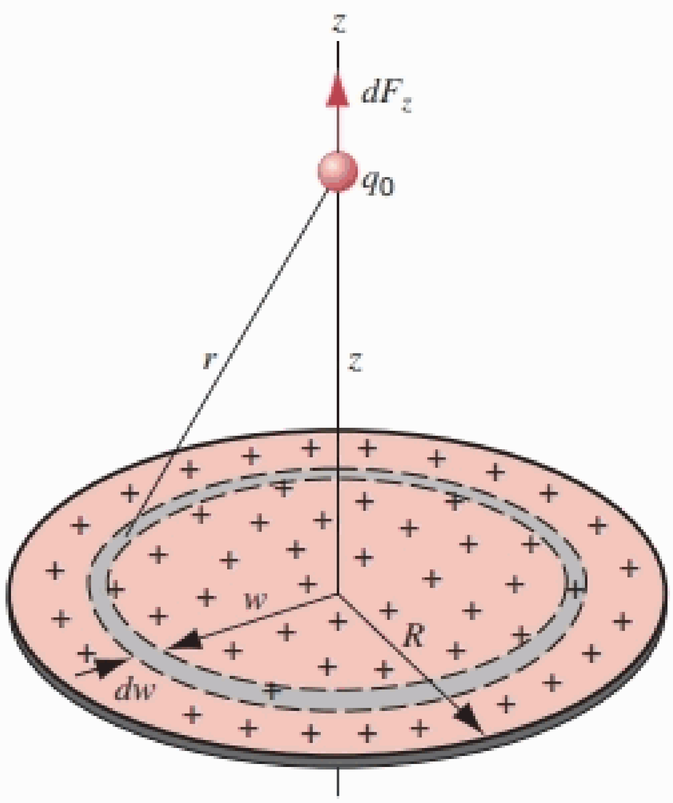


FIGURE 25-13. A circular disk carrying a uniform surface charge density. The force on a point charge q_0 is determined by dividing the disk into thin circular rings.

To add the force elements due to all of the rings, we integrate as w ranges from 0 to R:

$$\begin{aligned} F_z &= \frac{1}{4\pi\epsilon_0} q_0 2\pi\sigma z \int_0^R \frac{w\,dw}{(z^2+w^2)^{3/2}} \\ &= \frac{1}{4\pi\epsilon_0}\frac{2q_0 q}{R^2}\left(1 - \frac{z}{\sqrt{z^2+R^2}}\right). \qquad (25\text{-}17) \end{aligned}$$

Note that the integral is of the form $\int u^{-3/2}du$, which can be evaluated directly. How would this equation differ for $z < 0$? (See Exercise 15.) As $z \to \infty$, you can use the binomial expansion (see Appendix I) to show that this expression reduces to Coulomb's law for point charges.

In these three examples, we have assumed that all charges are positive. If either the point charge or the extended object (but not both!) carry negative charge, the direction of the force is opposite to what is shown in Figs. 25-11 to 25-13.

SAMPLE PROBLEM 25-6. Two circular disks of radius $R = 5.0$ cm are separated by 6.0 cm along a common vertical axis. The disks carry equal but opposite electric charges distributed uniformly over their surfaces. How much charge q must be placed on each disk to suspend a tiny drop of oil of mass 4.0×10^{-15} kg and charge $-e$ at a point along the axis of the disks and midway between them?

Solution Let the top disk be positively charged and the bottom disk be negatively charged. Then the top disk attracts the drop and the bottom disk repels it. Because the drop is midway between the disks, the forces have the same magnitude F_z given by Eq. 25-17. To hold the drop in equilibrium, the net upward electrostatic force $2F_z$ must equal the weight mg of the drop. Setting these two forces equal and using Eq. 25-17, we obtain

$$q = \frac{mg}{\dfrac{4e}{4\pi\epsilon_0 R^2}\left(1 - \dfrac{z}{\sqrt{z^2 + R^2}}\right)}.$$

Solving, we find $q = 35$ nC.

This method was used by Robert A. Millikan in a series of experiments begun in 1906 to measure the charge of the electron. (See Section 26-6.)

A Special Case

There is one special case in which a continuous charge distribution can be treated as a point charge, which allows us to use Coulomb's law in its point-charge form. This occurs when the charge is distributed with spherical symmetry. That is, the volume charge density may vary with radius, but the density is uniform in a thin shell at any radius.

First we consider a thin spherical shell. In Section 14-5 we established two properties of the gravitational force exerted by a uniformly dense spherical shell of matter on a point mass. These properties are: (1) the force on a particle inside the shell is zero, and (2) the force on an external particle is the same as if the entire mass of the shell were concentrated at its center.

The symmetry between the gravitational and electrostatic force laws (both depend on $1/r^2$) will allow us to make many analogies between gravitation and electrostatics. We can often apply results from gravitation directly to electrostatics without additional calculation or proof. This is true for the properties of uniform shells. The proofs of these two important results for electrostatics follow exactly the corresponding proofs in Section 14-5 for the gravitational force.

> *A uniformly charged spherical shell exerts no electrostatic force on a point charge located anywhere inside the shell.*

> *A uniformly charged spherical shell exerts an electrostatic force on a point charge outside the shell as if all the charge of the shell were concentrated in a point charge at its center.*

There is one difference between the gravitational and electrostatic cases: the gravitational force is always attractive, but the electrostatic force may be either attractive or repulsive. However, this difference does not affect the transfer of the above two rules from the gravitational to the electrostatic force.

These rules can be used to obtain a related result that is valid for spherical charge distributions. Suppose we have a spherical distribution of charge in which the volume charge density ρ is either constant or varies only as a function of the radius r. We can therefore regard the sphere as being composed of a set of thin spherical shells. Each shell is uniformly charged; that is, the charge density of one shell may differ from the charge density of another shell, but on each individual shell the charge is uniformly distributed. Then we can apply each of the two rules to every shell of the sphere. If our test charge is somewhere inside the sphere, the shells outside of the charge exert no force on it according to the first rule. (This result was used for the gravitational force in Sample Problem 14-4.) If the test charge is outside the sphere, then every shell can be replaced by a point charge at the center, and so the entire sphere can be replaced by a point charge equal to the total charge of the sphere.

For this reason the force exerted by the nucleus of an atom on its electrons generally cannot give us any information about the distribution of positive charge within the nucleus. For spherical nuclei in which the charge density depends only on r, *all* charge distributions give identical forces on an electron outside the nucleus. However, occasionally an electron may wander *inside* the nucleus and give us information about its distribution of positive charge.

SAMPLE PROBLEM 25-7. The spherical nucleus of a certain atom contains a positive charge Ze in a volume of radius R. Compare the force exerted on an electron inside the nucleus at radius $0.5R$ with the force at radius R for a nucleus in which (*a*) the charge density is constant throughout its volume, and (*b*) the charge density increases in direct proportion to the radius r.

Solution (*a*) For the uniformly charged nucleus (which is a good approximation to the behavior of many nuclei), the volume charge density is

$$\rho = \frac{q}{V} = \frac{Ze}{\frac{4}{3}\pi R^3}.$$

When the electron is at $r = R$, the entire nucleus can be replaced by a point charge $q = Ze$ located at its center, so the force $F(r)$ at $r = R$ is

$$F(R) = \frac{1}{4\pi\epsilon_0}\frac{Ze^2}{R^2}.$$

When the electron is at $r = R/2$, the charge at all larger radii exerts no force on the electron (by the first rule for spherical shells). The charge inside $R/2$ can be replaced with a point charge. How much charge is inside $R/2$? Since the volume of a sphere depends on r^3, the volume inside $r = R/2$ is $(1/2)^3 = 1/8$ the volume of the sphere. So the charge inside $r = R/2$ is 1/8 the charge of the entire sphere, and we conclude that

$$\frac{F(R/2)}{F(R)} = \frac{1}{8}.$$

(*b*) We can write the charge density as $\rho(r) = br$. First we must evaluate the constant of proportionality b. We know that the total charge on the nucleus must be Ze, so

$$\int_0^R \rho\, dV = \int_0^R (br)4\pi r^2\, dr = Ze,$$

where $dV = 4\pi r^2 dr$ is the volume of a spherical shell. Carrying out the integral, we find $b = Ze/\pi R^4$.

From the second rule for spherical shells, we know that the force $F(R)$ is the same for both charge distributions. However,

$F(R/2)$ will be different for the two distributions. To find $F(R/2)$, we need to know how much charge q' is contained within the sphere of radius $R/2$, since the charge outside of radius $R/2$ exerts no force on the electron. This charge is

$$q' = \int_0^{R/2} \rho\, dV = \int_0^{R/2} \frac{Zer}{\pi R^4} 4\pi r^2\, dr = \frac{Ze}{16}.$$

The force on the electron at $r = R/2$ can be found by replacing the sphere within $R/2$ by a point charge q' at its center, which gives a force 1/16 the force at the surface:

$$\frac{F(R/2)}{F(R)} = \frac{1}{16}.$$

This result is very different from that of the uniform sphere in part (*a*), demonstrating that, although the electron outside the nucleus cannot distinguish between the two distributions, the electron inside clearly can.

Atomic electrons can occasionally penetrate the nucleus, and electrons can be fired into the nucleus by an accelerator. These two methods can give us information about the distribution of charge within the nucleus. One result of these experiments is that the charge density is found to be very nearly uniform for most nuclei. Despite the Coulomb repulsion of the protons (which is expected to drive them toward the nuclear surface) and despite the strong nuclear force between the protons (which is expected to make them congregate near the center of the nucleus), the protons in the nucleus are distributed with a roughly uniform density. Furthermore, this density is approximately the same for light nuclei as it is for heavy nuclei. These surprising results offer insight into important properties of the nuclear force.

25-6 CONSERVATION OF CHARGE

When a glass rod is rubbed with silk, a positive charge appears on the rod. Measurement shows that a corresponding negative charge appears on the silk. This suggests that rubbing does not create charge but merely transfers it from one object to another, disturbing slightly the electrical neutrality of each. This hypothesis of the *conservation of charge* has stood up under careful experimental tests both for large-scale objects and for atoms, nuclei, and elementary particles. No exceptions have ever been found.

In analogy with other conservation laws, such as conservation of momentum or conservation of energy, we can express conservation of electric charge as

$$\sum q = \text{constant} \quad \text{or} \quad q_\mathrm{i} = q_\mathrm{f}. \tag{25-18}$$

In any process occurring in an isolated system the net initial charge must equal the net final charge. In finding the net charge, it is important to take into account the signs of the individual charges.

An interesting example of charge conservation comes about when an electron (charge $= -e$) and an antielectron or positron (charge $= +e$) are brought close to each other. The two particles may annihilate one another, converting all their rest energy into radiant energy. The radiant energy may appear in the form of two gamma rays (high-energy packets of electromagnetic radiation, which are chargeless):

$$\mathrm{e}^- + \mathrm{e}^+ \rightarrow \gamma + \gamma.$$

The net charge is zero both before and after the event, and charge is conserved.

Certain uncharged particles, such as the neutral π meson, sometimes decay into two gamma rays:

$$\pi^0 \rightarrow \gamma + \gamma.$$

This decay conserves charge, the total charge again being 0 before and after the decay. For another example, a neutron ($q = 0$) decays into a proton ($q = +e$) and an electron ($q = -e$) plus another neutral particle, a neutrino ($q = 0$):

$$\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e}^- + \nu.$$

The total charge is zero, both before and after the decay, and charge is conserved. Experiments have been done to search for decays of the neutron into a proton with no electron emitted, which would violate charge conservation. No such events have been found.

The decay of an electron ($q = -e$) into neutral particles, such as gamma rays (γ) or neutrinos (ν), is forbidden; for example

$$\mathrm{e}^- \nrightarrow \gamma + \nu,$$

because that decay would violate charge conservation. Attempts to observe this decay have likewise been unsuccessful, indicating that, if the decay does occur, the electron must have a lifetime of at least 10^{23} years!

Another example of charge conservation is found in the fusion of two deuterium nuclei $^2\mathrm{H}$ (called "heavy hydrogen") to make helium. Among the possible reactions are

$$^2\mathrm{H} + {}^2\mathrm{H} \rightarrow {}^3\mathrm{H} + \mathrm{p},$$

$$^2\mathrm{H} + {}^2\mathrm{H} \rightarrow {}^3\mathrm{He} + \mathrm{n}.$$

The deuterium nucleus contains one proton and one neutron and therefore has a charge of $+e$. The nucleus of the isotope of hydrogen with mass 3, written $^3\mathrm{H}$ and known as *tritium,* contains one proton and two neutrons, and thus also has a charge of $+e$. The first reaction therefore has a net charge of $+2e$ on each side and conserves charge. In the second reaction, the neutron is uncharged, while the nucleus of the isotope of helium with mass 3 contains two protons and one neutron and therefore has a charge of $+2e$. The second reaction thus also conserves charge. Conservation of charge explains why we never see a proton emitted along with $^3\mathrm{He}$ or a neutron along with $^3\mathrm{H}$.

To sum up, charge is conserved in *all* known interactions among particles; no exceptions have ever been observed.

MULTIPLE CHOICE

25-1 Electromagnetism: A Preview

25-2 Electric Charge

1. Electric charges A and B are attracted to each other. Electric charges B and C repel each other. If A and C are held close together they will

(A) attract. (B) repel.
(C) not affect each other.
(D) More information is needed to answer.

2. Electric charges A and B are attracted to each other. Electric charges B and C are also attracted to each other. If A and C are held close together they will

(A) attract. (B) repel.
(C) not affect each other.
(D) More information is needed to answer.

3. Electric charges A and B repel each other. Electric charges B and C also repel each other. If A and C are held close together they will

(A) attract. (B) repel.
(C) not affect each other.
(D) More information is needed to answer.

25-3 Conductors and Insulators

4. If an object made of substance A rubs an object made of substance B, then A becomes positively charged and B becomes negatively charged. If, however, an object made of substance A is rubbed against an object made of substance C, then A becomes negatively charged. What will happen if an object made of substance B is rubbed against an object made of substance C?

(A) B becomes positively charged and C becomes positively charged.
(B) B becomes positively charged and C becomes negatively charged.
(C) B becomes negatively charged and C becomes positively charged.
(D) B becomes negatively charged and C becomes negatively charged.

5. A positively charged rod is held near a ball suspended by an insulating thread. The ball is seen to swing toward the charged rod. What can be concluded?

(A) The ball must have had a charge opposite to that of the rod.
(B) The ball may have been neutral originally, but it became charged when the rod was held near it.
(C) The ball must be a conductor.
(D) The ball is not positively charged, but it could be neutral.

6. A spherical conducting ball is suspended by a grounded conducting thread. A positive point charge is moved near the ball. The ball will

(A) be attracted to the point charge and swing toward it.
(B) be repelled from the point charge and swing away from it.
(C) not be affected by the point charge.

7. A spherical conducting ball is suspended by an insulating thread. A positive point charge is moved near the ball. The ball will

(A) be attracted to the point charge and swing toward it.
(B) be repelled from the point charge and swing away from it.
(C) not be affected by the point charge.

25-4 Coulomb's Law

8. A 3-μC point charge q_1 is located a distance d away from a $-$6-μC point charge q_2. What is the ratio $|\vec{\mathbf{F}}_{12}|/|\vec{\mathbf{F}}_{21}|$?

(A) 1/2 (B) 1 (C) 2 (D) 18

9. Two 200-pound lead balls are separated by a distance of 1 m. Both balls have the same positive charge q. What charge will produce an electrostatic force between the balls that is of the same order of magnitude as the weight of one ball?

(A) 1×10^{-14} C (B) 1×10^{-7} C
(C) 3×10^{-4} C (D) 2×10^{-2} C

10. Two identical, small, conducting spheres are separated by a distance of 1 m. The spheres originally have the same positive charge, and the force between them is F_0. Half of the charge on one sphere is then moved to the other sphere. The force between the spheres is now

(A) $F_0/4$. (B) $F_0/2$. (C) $3F_0/4$.
(D) $3F_0/2$. (E) $3F_0$.

11. Two identical, small, conducting spheres are separated by a distance of 1 m. The spheres originally have equal but opposite charges, and the force between them is F_0. Half of the charge on one sphere is then moved to the other sphere. The force between the spheres is now

(A) $F_0/4$. (B) $F_0/2$. (C) $3F_0/4$.
(D) $3F_0/2$. (E) $3F_0$.

25-5 Continuous Charge Distributions

12. A point charge q is located a distance a from the surface of a sphere of radius $2a$. A charge Q is distributed uniformly throughout the volume of the sphere. The magnitude of the electrostatic force between the point charge q and the sphere is F, where

(A) $F = |qQ|/4\pi\epsilon_0 a^2$.
(B) $|qQ|/4\pi\epsilon_0 a^2 > F > |qQ|/12\pi\epsilon_0 a^2$.
(C) $|qQ|/12\pi\epsilon_0 a^2 > F > |qQ|/20\pi\epsilon_0 a^2$.
(D) $|qQ|/20\pi\epsilon_0 a^2 > F > |qQ|/36\pi\epsilon_0 a^2$.
(E) $F = |qQ|/36\pi\epsilon_0 a^2$.

25-6 Conservation of Change

13. A positively charged rod is held near a neutral conducting sphere suspended by an insulating thread. The sphere will

(A) be unaffected, because it is neutral.
(B) remain neutral, but be repelled from the rod anyway.
(C) remain neutral, but be attracted to the rod anyway.
(D) acquire a negative charge and be repelled from the rod.
(E) acquire a negative charge and be attracted to the rod.

14. Objects A, B, and C are three identical, insulated, spherical conductors. Originally A and B both have charges of $+3$ mC, while C has a charge of -6 mC. Objects A and C are allowed to touch, then they are moved apart. Then objects B and C are allowed to touch, and they are moved apart.

(*a*) If objects A and B are now held near each other, they will
(A) attract. (B) repel.
(C) have no effect on each other.

(*b*) If instead objects A and C are held near each other, they will
(A) attract. (B) repel.
(C) have no effect on each other.

QUESTIONS

1. You are given two metal spheres mounted on portable insulating supports. Find a way to give them equal and opposite charges. You may use a glass rod rubbed with silk but may not touch it to the spheres. Do the spheres have to be of equal size for your method to work?

2. In Question 1, find a way to give the spheres equal charges of the same sign. Again, do the spheres need to be of equal size for your method to work?

3. A charged rod attracts bits of dry cork dust, which, after touching the rod, often jump violently away from it. Explain.

4. How would your answers to Multiple-choice question 1, Multiple-choice question 2, and Multiple-choice question 3 change if any of the objects *A*, *B*, or *C* could be uncharged?

5. The experiments described in Section 25-2 could be explained by postulating four kinds of charge—that is, on glass, silk, plastic, and fur. What is the argument against this?

6. A positive charge is brought very near to an uncharged insulated conductor. The conductor is grounded while the charge is kept near. Is the conductor charged positively, negatively, or not at all if (*a*) the charge is taken away and then the ground connection is removed and (*b*) the ground connection is removed and then the charge is taken away?

7. A charged insulator can be discharged by passing it just above a flame. Explain how.

8. If you rub a coin briskly between your fingers, it will not seem to become charged by friction. Why?

9. If you walk briskly across a carpet, you often experience a spark on touching a door knob. (*a*) What causes this? (*b*) How might it be prevented?

10. Why do electrostatic experiments not work well on humid days?

11. Why is it recommended that you touch the metal frame of your personal computer before installing any internal accessories?

12. An insulated rod is said to carry an electric charge. How could you verify this and determine the sign of the charge?

13. If a charged glass rod is held near one end of an insulated uncharged metal rod as in Fig. 25-14, electrons are drawn to one end, as shown. Why does the flow of electrons cease? After all, there is an almost inexhaustible supply of them in the metal rod.

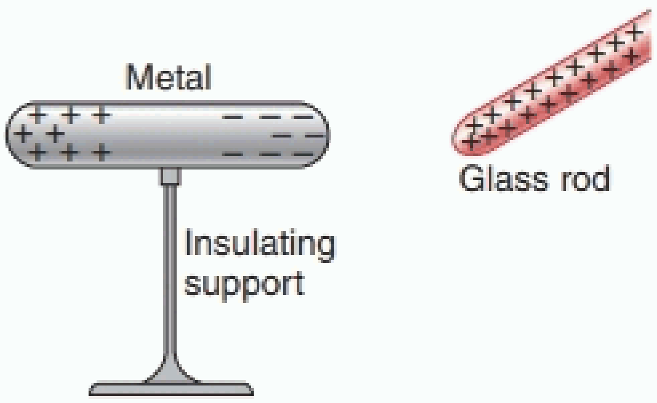


FIGURE 25-14. Questions 13 and 14

14. In Fig. 25-14, does any resultant electric force act on the metal rod? Explain.

15. A person standing on an insulating stool touches a charged, insulated conductor. Is the conductor discharged completely?

16. (*a*) A positively charged glass rod attracts a suspended object. Can we conclude that the object is negatively charged? (*b*) A positively charged glass rod repels a suspended object. Can we conclude that the object is positively charged?

17. Explain what is meant by the statement that electrostatic forces obey the principle of superposition.

18. Is the electric force that one charge exerts on another changed if other charges are brought nearby?

19. A solution of copper sulfate is a conductor. What particles serve as the charge carriers in this case?

20. If the electrons in a metal such as copper are free to move about, they must often find themselves headed toward the metal surface. Why do they not keep on going and leave the metal?

21. Would it have made any important difference if Benjamin Franklin had chosen, in effect, to call electrons positive and protons negative?

22. Coulomb's law predicts that the force exerted by one point charge on another is proportional to the product of the two charges. How might you go about testing this aspect of the law in the laboratory?

23. Explain how an atomic nucleus can be stable if it is composed of particles that are either neutral (neutrons) or carry like charges (protons).

24. An electron (charge $= -e$) circulates around a helium nucleus (charge $= +2e$) in a helium atom. Which particle exerts the larger force on the other?

25. The charge of a particle is a true characteristic of the particle, independent of its state of motion. Explain how you can test this statement by making a vigorous experimental check of whether the hydrogen atom is truly electrically neutral.

26. Suppose the charge in Fig. 25-11 were not distributed uniformly along the length of the rod but instead were concentrated at its center and tapered off at the same rate toward either end. Will the force now have a z component? If this rod had the same total charge q as the uniformly charged rod, how will the magnitude of F_y compare with Eq. 25-15? Repeat both questions if the charge is distributed along the rod so that there is a deficiency near the center and the charge density increases at the same rate toward either end.

27. Earnshaw's theorem says that no particle can be in stable equilibrium under the action of electrostatic forces alone. Consider, however, point P at the center of a square of four equal positive charges, as in Fig. 25-15. If you put a positive test charge there it might seem to be in stable equilibrium. Every one of the four external charges pushes it toward P, yet Earnshaw's theorem holds. Can you explain how?

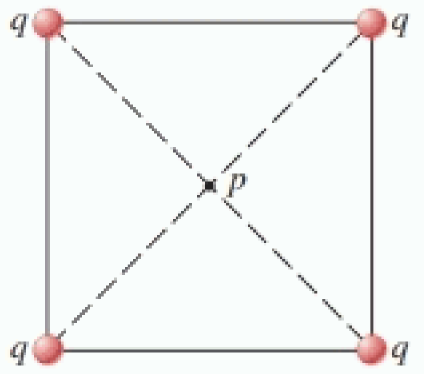


FIGURE 25-15. Question 27.

28. The quantum of charge is 1.60×10^{-19} C. Is there a corresponding quantum of mass?

29. What does it mean to say that a physical quantity is (*a*) quantized or (*b*) conserved? Give some examples.

30. In Sample Problem 25-3 we show that the electrical force is about 10^{39} times stronger than the gravitational force. Can you conclude from this that a galaxy, a star, or a planet must be essentially neutral electrically?

31. How do we know that electrostatic forces are not the cause of gravitational attraction—between the Earth and Moon, for example?

EXERCISES

25-1 Electromagnetism: A Preview

25-2 Electric Charge

1. In the return stroke of a typical lightning bolt (see Fig. 25-16), a current of 2.5×10^4 C/s flows for 20 μs. How much charge is transferred in this event?

FIGURE 25-16. Exercise 1.

25-3 Conductors and Insulators

25-4 Coulomb's Law

2. What must be the distance between point charge $q_1 = 26.3\ \mu$C and point charge $q_2 = -47.1\ \mu$C for the attractive electrical force between them to have a magnitude of 5.66 N?

3. A point charge of $+3.12 \times 10^{-6}$ C is 12.3 cm distant from a second point charge of -1.48×10^{-6} C. Calculate the magnitude of the force on each charge.

4. Two equally charged particles, held 3.20 mm apart, are released from rest. The initial acceleration of the first particle is observed to be 7.22 m/s^2 and that of the second to be 9.16 m/s^2. The mass of the first particle is 6.31×10^{-7} kg. Find (*a*) the mass of the second particle and (*b*) the magnitude of the common charge.

5. Figure 25-17*a* shows two charges, q_1 and q_2, held a fixed distance d apart. (*a*) Find the strength of the electric force that acts on q_1. Assume that $q_1 = q_2 = 21.3\ \mu$C and $d = 1.52$ m. (*b*) A third charge $q_3 = 21.3\ \mu$C is brought in and placed as shown in Fig. 25-17*b*. Find the strength of the electric force on q_1 now.

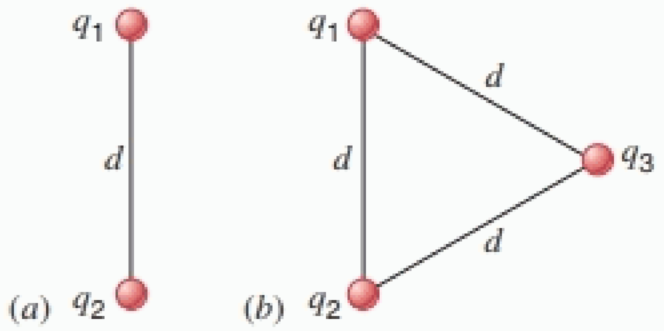


FIGURE 25-17. Exercise 5.

6. Two identical conducting spheres, 1 and 2, carry equal amounts of charge and are fixed a distance apart that is large compared with their diameters. The spheres repel each other with an electrical force of 88 mN. Suppose now that a third identical sphere 3, having an insulating handle and initially uncharged, is touched first to sphere 1, then to sphere 2, and finally removed. Find the force between spheres 1 and 2 now. See Fig. 25-18.

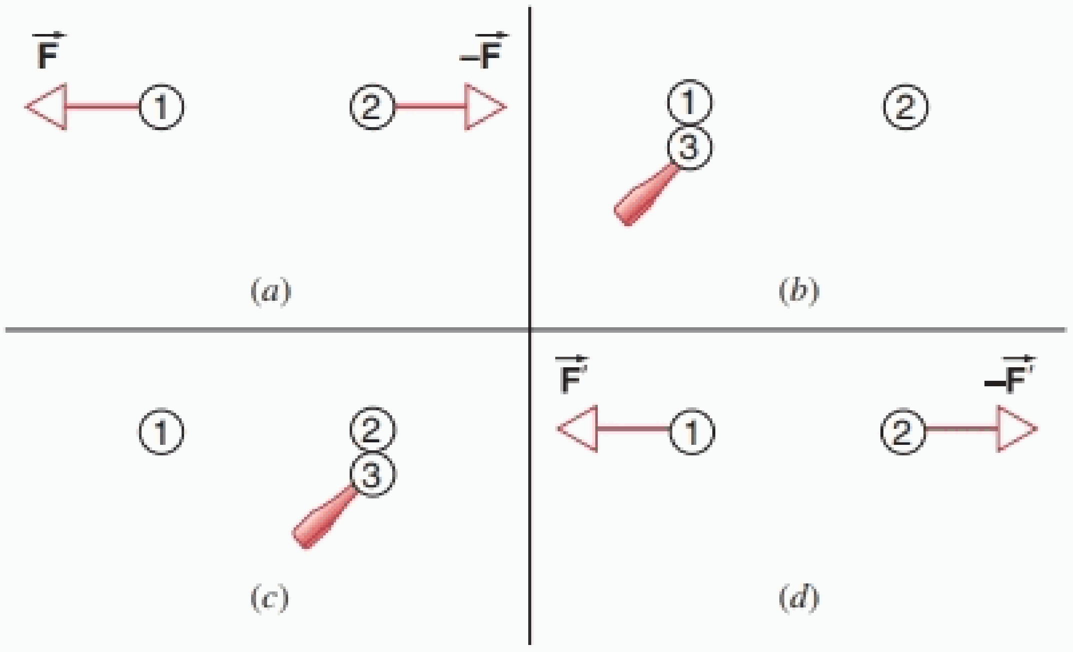


FIGURE 25-18. Exercise 6.

7. Three charged particles lie on a straight line and are separated by a distance d as shown in Fig. 25-19. Charges q_1 and q_2 are held fixed. Charge q_3, which is free to move, is found to be in equilibrium under the action of the electric forces. Find q_1 in terms of q_2.

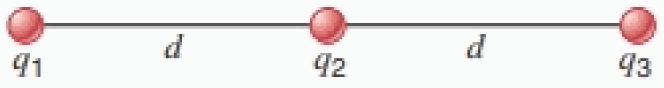


FIGURE 25-19. Exercise 7.

8. In Fig. 25-20, find (*a*) the horizontal components and (*b*) the vertical components of the resultant electric force on the charge in the lower left corner of the square. Assume that $q = 1.13\ \mu\text{C}$ and $a = 15.2$ cm. The charges are at rest.

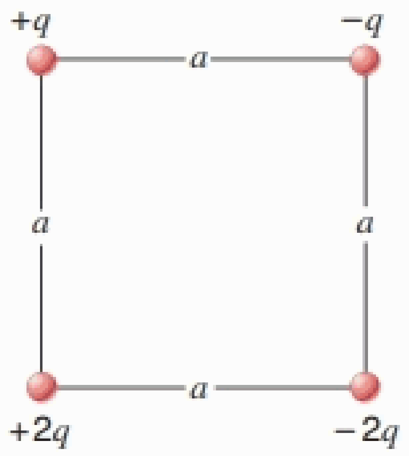


FIGURE 25-20. Exercise 8.

9. Two positive charges, each of $4.18\ \mu\text{C}$, and a negative charge, $-6.36\ \mu\text{C}$, are fixed at the vertices of an equilateral triangle of side 13.0 cm. Find the electrical force on the negative charge.

10. Each of two small spheres is charged positively, the total charge being $52.6\ \mu\text{C}$. Each sphere is repelled from the other with a force of 1.19 N when the spheres are 1.94 m apart. Calculate the charge on each sphere.

11. Two fixed charges, $+1.07\ \mu\text{C}$ and $-3.28\ \mu\text{C}$, are 61.8 cm apart. Where may a third charge be located so that no net force acts on it?

12. Three small balls, each of mass 13.3 g, are suspended separately from a common point by silk threads, each 1.17 m long. The balls are identically charged and hang at the corners of an equilateral triangle 15.3 cm on a side. Find the charge on each ball.

13. A cube of edge a carries a point charge q at each corner. Show that the resultant electric force on any one of the charges is given by

$$F = \frac{0.262q^2}{\epsilon_0 a^2},$$

directed along the body diagonal away from the cube.

25-5 Continuous Charge Distributions

14. Equation 25-15 was derived assuming the charge q_0 was located on the positive y axis. (*a*) Does Eq. 25-15 remain valid if the charge is located at negative y? Explain. (*b*) Write an equation similar to Eq. 25-15 if the point charge q_0 is instead located a distance x from the rod on the positive or negative x axis. (*c*) Write an equation in vector component form for the force when q_0 is located a distance d from the rod on the 45° line that bisects the positive x and y axes. (*d*) Write an equation in vector component form that gives the force when q_0 is located at an arbitrary point x,y anywhere in the xy plane. Check that the components have the correct signs when the point x,y is located in each of the four quadrants.

15. (*a*) Starting with Eq. 25-16, write an equation in vector form that gives the force when q_0 is located either on the positive or negative z axis of the ring of charge. (*b*) Do the same for the disk of charge using Eq. 25-17.

16. Find the force on a positive point charge q located a distance x from the end of a rod of length L with uniformly distributed positive charge Q. (See Fig. 25-21.)

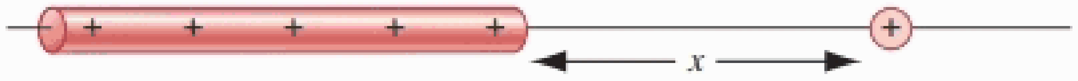


FIGURE 25-21. Exercise 16.

17. Consider the rod and charge q_0 in Fig. 25-11. Where can you place a second point charge q (equal to the charge on the rod) so that q_0 is in equilibrium (ignore gravity)? Solve this problem assuming that (*a*) q is positive and (*b*) q is negative.

18. Show that the equilibrium of q_0 in Exercise 17 is unstable. (Hint: This problem can be solved by symmetry arguments, and actually requires very little math!)

19. Assume that the rod in Fig. 25-11 has a uniform positive charge density λ on the top half of the rod and a uniform charge density $-\lambda$ on the bottom half of the rod. Find the net force on the point charge q_0.

20. Four charged rods form the sides of a square in the horizontal (xy) plane. Each rod has a length $L = 25.0$ cm and each carries a uniformly distributed positive charge Q. A small sphere, which can be considered to be a point charge of mass 3.46×10^{-4} g and charge $q = +2.45 \times 10^{-12}$ C, is in equilibrium a distance $z = 21.4$ cm above the center of the square. Find the value of Q.

25-6 Conservation of Charge

21. Identify the element X in the following nuclear reactions:
(*a*) ${}^{1}\text{H} + {}^{9}\text{Be} \rightarrow \text{X} + \text{n}$;
(*b*) ${}^{12}\text{C} + {}^{1}\text{H} \rightarrow \text{X}$;
(*c*) ${}^{15}\text{N} + {}^{1}\text{H} \rightarrow {}^{4}\text{He} + \text{X}$.
(Hint: See Appendix E.)

22. In the radioactive decay of ${}^{238}\text{U}$ (${}^{238}\text{U} \rightarrow {}^{4}\text{He} + {}^{234}\text{Th}$), the center of the emerging ${}^{4}\text{He}$ particle is, at a certain instant, 12×10^{-15} m from the center of the residual ${}^{234}\text{Th}$ nucleus. At this instant, (*a*) what is the force on the ${}^{4}\text{He}$ particle, and (*b*) what is its acceleration?

23. In a crystal of salt, an atom of sodium transfers one of its electrons to a neighboring atom of chlorine, forming an ionic bond. The resulting positive sodium ion and negative chlorine ion attract each other by the electrostatic force. Calculate the force of attraction if the ions are 282 pm apart.

24. The electrostatic force between two identical ions that are separated by a distance of 5.0×10^{-10} m is 3.7×10^{-9} N. (*a*) Find the charge on each ion. (*b*) How many electrons are missing from each ion?

25. A neutron is thought to be composed of one "up" quark of charge $+\frac{2}{3}e$ and two "down" quarks each having charge $-\frac{1}{3}e$. If the down quarks are 2.6×10^{-15} m apart inside the neutron, what is the repulsive electrical force between them?

26. (*a*) How many electrons would have to be removed from a penny to leave it with a charge of $+1.15 \times 10^{-7}$ C? (*b*) To what fraction of the electrons in the penny does this correspond? See Sample Problem 25-1.

27. An electron is in a vacuum near the surface of the Earth. Where should a second electron be placed so that the net force on the first electron, due to the other electron and to gravity, is zero?

28. Find the total charge in coulombs of 75.0 kg of electrons.

29. Calculate the number of coulombs of positive charge in a glass of water. Assume the volume of the water to be 250 cm^3.

30. Two physics students (Mary at 52.0 kg and John at 90.7 kg) are 28.0 m apart. Let each have a 0.01% imbalance in their amounts of positive and negative charge, one student being positive and the other negative. Estimate the electrostatic force of attraction between them. (Hint: Replace the students by spheres of water and use the result of Exercise 29.)

31. (*a*) What equal amounts of positive charge would have to be placed on the Earth and on the Moon to neutralize their gravitational attraction? Do you need to know the Moon's distance to solve this problem? Why or why not? (*b*) How many metric tons of hydrogen would be needed to provide the positive charge calculated in part (*a*)? The molar mass of hydrogen is 1.008 g/mol.

PROBLEMS

1. Two identical conducting spheres, having charges of opposite sign, attract each other with a force of 0.108 N when separated by 50.0 cm. The spheres are suddenly connected by a thin conducting wire, which is then removed, and thereafter the spheres repel each other with a force of 0.0360 N. What were the initial charges on the spheres?

2. A charge Q is fixed at each of two opposite corners of a square. A charge q is placed at each of the other two corners. (*a*) If the resultant electrical force on Q is zero, how are Q and q related? (*b*) Could q be chosen to make the resultant electrical force on every charge zero? Explain your answer.

3. Two free point charges $+q$ and $+4q$ are a distance L apart. A third charge is placed so that the entire system is in equilibrium. (*a*) Find the sign, magnitude, and location of the third charge. (*b*) Show that the equilibrium is unstable.

4. Two similar tiny balls of mass m are hung from silk threads of length L and carry equal charges q as in Fig. 25-22. Assume that θ is so small that $\tan\theta$ can be replaced by its approximate equal, $\sin\theta$. (*a*) To this approximation show that, for equilibrium,

$$x = \left(\frac{q^2L}{2\pi\epsilon_0 mg}\right)^{1/3},$$

where x is the separation between the balls. (*b*) If $L =$ 122 cm, $m = 11.2$ g, and $x = 4.70$ cm, what is the value of q?

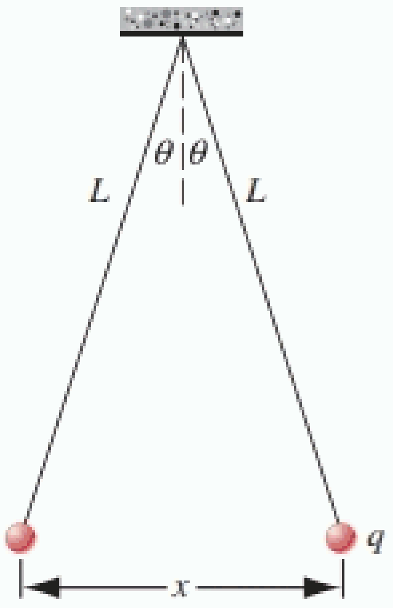


FIGURE 25-22. Problems 4, 5, and 6.

5. If the balls of Fig. 25-22 are conducting, (*a*) what happens to them after one is discharged? Explain your answer. (*b*) Find the new equilibrium separation.

6. Assume that each ball in Problem 4 is losing charge at the rate of 1.20 nC/s. At what instantaneous relative speed ($= dx/dt$) do the balls approach each other initially?

7. A certain charge Q is to be divided into two parts, $Q - q$ and q. What is the relation of Q to q if the two parts, placed a given distance apart, are to have a maximum Coulomb repulsion?

8. Two positive charges $+Q$ are held fixed a distance d apart. A particle of negative charge $-q$ and mass m is placed midway between them, then is given a small displacement perpendicular to the line joining them and released. Show that the particle describes simple harmonic motion of period $(\epsilon_0 m\pi^3 d^3/qQ)^{1/2}$.

9. Calculate the period of oscillation for a particle of positive charge $+q$ displaced from the midpoint and along the line joining the two charges $+Q$ in Problem 8.

10. In the compound CsCl (cesium chloride), the Cs atoms are situated at the corners of a cube with a Cl atom at the cube's center. The edge length of the cube is 0.40 nm; see Fig. 25-23. The Cs atoms are each deficient in one electron and the Cl atom carries one excess electron. (*a*) What is the strength of the net electric force on the Cl atom resulting from the eight Cs atoms shown? (*b*) Suppose that the Cs atom marked with an arrow is missing (crystal defect). What now is the net electric force on the Cl atom resulting from the seven remaining Cs atoms?

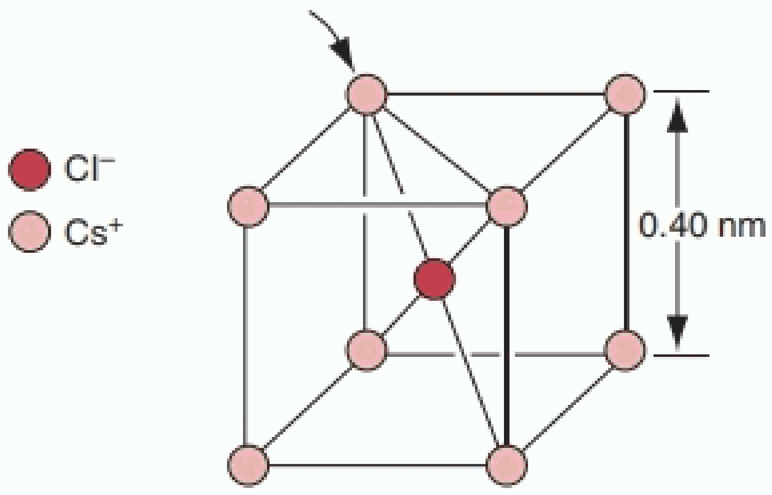


FIGURE 25-23. Problem 10.

11. Two equal positive point charges q are held a fixed distance $2a$ apart. A point test charge is located in a plane that is normal to the line joining these charges and midway between them. Find the radius R of the circle in this plane for which the force on the test particle has a maximum value. See Fig. 25-24.

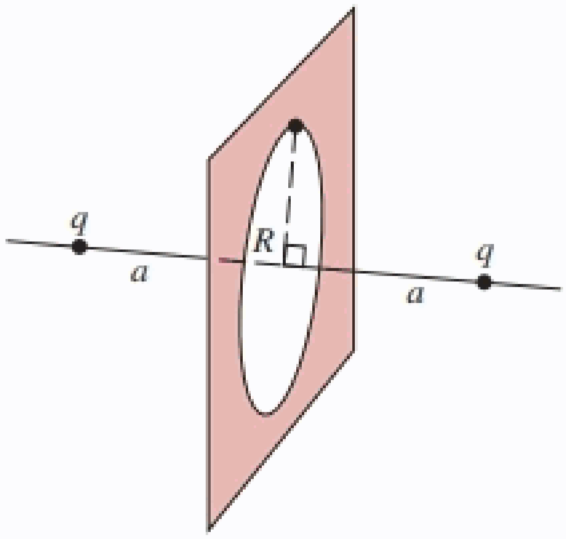


FIGURE 25-24. Problem 11.

COMPUTER PROBLEMS

1. Calculate the force of attraction between two rings with uniformly distributed charges $+q$ and $-q$. The axis of the rings is the x axis, each has a radius R, and the rings are separated by a distance of $2R$. Your final answer should be of the form $F = C_r q^2/4\pi\epsilon_0 R^2$, where C_r is a dimensionless constant that you will find.

2. Repeat Computer Problem 1 for the case of two disks with uniformly distributed charges $+q$ and $-q$. Your final answer will still be of the form $F = C_d q^2/4\pi\epsilon_0 R^2$, where C_d is a dimensionless constant that you will find; C_d for the disk *is different* from C_r for the ring.

3. Calculate the force of attraction between two solid spheres with uniformly distributed charges $+q$ and $-q$. The spheres are centered on the x axis, each has a radius R, and the *centers* of the spheres are separated by a distance of $d > 2R$. Your final answer should be of the form $F = C_s q^2/4\pi\epsilon_0 d^2$, where C_s is a dimensionless constant that you will find.

4. A uniform ring of charge Q has a radius $R = 1.00$ cm. An electron is constrained to move in the plane of the ring. (a) Assuming that $Q = -100\ \mu\text{C}$, find the speed of an electron that would move in a circular orbit of radius $r = 0.50$ cm, concentric with the ring. (b) Assuming that $Q = +100\ \mu\text{C}$, find the speed of an electron that would move in a circular orbit of radius $r = 1.50$ cm, concentric with the ring. (c) Show, by numerically integrating the motion, that neither orbit is stable.

CHAPTER 26

THE ELECTRIC FIELD

Electric charges can interact with one another over vast distances. Electrons or ionized atoms at the furthest reaches of the known universe can exert forces that cause electrons to move on Earth.

How can we explain these interactions? We do so in terms of the electric field—the distant charges set up an electric field, which exists throughout the space between Earth and the origin of the field. The motion of the charges causes disturbances in the field, which travel through space with the speed of light and are detected eons later (as radiation) when they cause motion of electrons in circuits on Earth.

In this chapter we consider only the static electric field due to charges at rest. Later in this book we expand our discussion to show how moving charges are responsible for the fields associated with electromagnetic radiation, such as radio or light.

26-1 WHAT IS A FIELD?

The temperature has a definite value at every point in the room in which you are sitting. You can measure the temperature at each point by putting a thermometer there. You could then represent that temperature distribution either by drawing a map of the room showing the measured temperature at each point, or else by specifying a mathematical function $T(x, y, z)$ that can be used to calculate the temperature at any point x, y, z. This distribution of temperature, represented either as a map or a function, is called a *temperature field.* In a similar way we might measure the pressure at each point and so determine the *pressure field.* These two fields are examples of *scalar fields,* because the temperature and the pressure are scalar quantities. If the temperature or pressure do not vary with time, they are also *static fields;* otherwise they are *time-varying fields* that might be represented mathematically by a time-dependent function such as $T(x, y, z, t)$.

If, on the other hand, you wanted to measure the velocity at every point in a flowing fluid, you would need to specify the value of the velocity *vector* at each point. Once again, you could draw a map showing the magnitude and direction of the velocity at any point, or you could specify a mathematical function $\vec{\mathbf{v}}(x, y, z)$ that would allow the flow velocity to be calculated at any point. This is an example of a *vector field.*

The Earth's gravitational field is another example of a vector field. You might measure the value of the gravitational force at any point by attaching a test mass m_0 to a spring scale. You could then determine the magnitude and direction of the gravitational force $\vec{\mathbf{F}}$ at any point, and you could present your results either by drawing a map showing the magnitude and direction of the force at various points or by specifying a mathematical function $\vec{\mathbf{F}}(x, y, z)$. However, such a map would not be useful to other people unless they used exactly the same test mass that you used. Since the force you measure is directly proportional to the value of the test mass, a better procedure would be to produce a map showing not the force on your test mass but instead the force per unit test mass, or $\vec{\mathbf{F}}/m_0$. This quantity, which would have units of N/kg, would be independent of the value of the test mass m_0. Choosing a test mass of a different size would give exactly the same map with the same values of the force per unit mass at every point.* We call the quantity $\vec{\mathbf{F}}/m_0$ the *gravitational field.* You will recognize it as also being equal to the free-fall acceleration $\vec{\mathbf{g}}$ at any point:

$$\vec{\mathbf{g}} = \frac{\vec{\mathbf{F}}}{m_0}. \tag{26-1}$$

The field $\vec{\mathbf{g}}$ is a vector whose direction gives the direction of the gravitational force at that point and whose magnitude indicates the "strength" of the gravitational effect at that point. We could find the force on a mass m at any point by multiplying $\vec{\mathbf{g}}$ at that point by the value of the mass:

$$\vec{\mathbf{F}} = m\vec{\mathbf{g}}. \tag{26-2}$$

In this chapter, we will develop the useful concept of an *electric field* based on a similar procedure that involves determining the electrical force per unit charge (rather than the gravitational force per unit mass). Because a force is involved, the electric field is a vector field. For now we deal only with static fields, but later when we discuss electromagnetic radiation we will consider time-varying electric fields.

Before the concept of fields became widely accepted, the force exerted by one gravitating body on another was thought of as a direct and instantaneous interaction. This view, called *action at a distance,* was also used for electromagnetic forces. In the case of gravitation, it can be represented schematically as

$$\text{mass} \rightleftarrows \text{mass},$$

indicating that the two masses interact directly with one another. According to this view, the effect of a movement of one body is instantaneously transmitted to the other body. This view violates the special theory of relativity, which limits the speed at which such information can be transmitted to the speed of light c, at most. The present interpretation, based on the field concept, can be represented as

$$\text{mass} \rightleftarrows \text{field} \rightleftarrows \text{mass},$$

in which each mass interacts not directly with the other but instead with the gravitational field established by the other. That is, the first mass sets up a field that has a certain value at every point in space; the second mass then interacts with the field at its particular location. The field plays the role of an intermediary between the two bodies. The force exerted on the second mass can be calculated from Eq. 26-2, given the value of the field $\vec{\mathbf{g}}$ due to the first mass. The situation is completely symmetrical from the point of view of the first mass, which interacts with the gravitational field established by the second mass. Changes in the location of one mass cause variations in its gravitational field; these variations travel at the speed of light, so the field concept is consistent with the restrictions imposed by special relativity.

26-2 THE ELECTRIC FIELD

The previous description of the gravitational field can be carried directly over to electrostatics. Coulomb's law for the force of one electric charge on another encourages us to think in terms of action at a distance, represented as

$$\text{charge} \rightleftarrows \text{charge}.$$

Again introducing the field as an intermediary between the charges, we can represent the interaction as

$$\text{charge} \rightleftarrows \text{field} \rightleftarrows \text{charge}.$$

That is, the first charge sets up an *electric field,* and the second charge interacts with the electric field of the first charge. Our problem of determining the interaction between the charges is therefore reduced to two separate problems: (1) determine, by measurement or calculation, the electric field established by the first charge at every point in space, and (2) calculate the force that the field exerts on the second charge placed at a particular point in space.

In analogy with Eq. 26-1 for the gravitational field, we define the electric field $\vec{\mathbf{E}}$ associated with a certain collection of charges in terms of the force exerted on a positive test charge q_0 at a particular point, or

$$\vec{\mathbf{E}} = \frac{\vec{\mathbf{F}}}{q_0}. \tag{26-3}$$

The direction of the vector $\vec{\mathbf{E}}$ is the same as the direction of $\vec{\mathbf{F}}$, because q_0 is a positive scalar. Defined in this way, the electric field is independent of the magnitude of the test charge q_0.

Figure 26-1 suggests how we use this definition to determine the electric field at a particular point P. We place a positive test charge q_0 at P, and we determine the electrostatic force on q_0 due to objects in the surrounding area, which are not shown in the figure. Equation 26-3 then gives the electric field shown in Fig. 26-1b. Note that $\vec{\mathbf{E}}$ and $\vec{\mathbf{F}}$ are parallel, as they must be from the definition of Eq. 26-3.

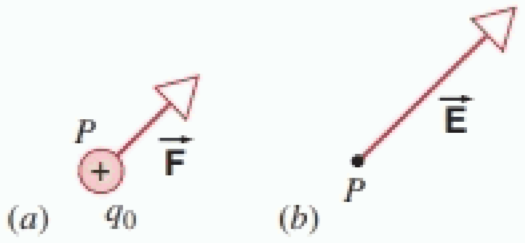


FIGURE 26-1. (a) Charged objects in the environment exert a force $\vec{\mathbf{F}}$ on a positive test charge q_0 at point P. (b) The electric field at point P due to the charged objects in the environment.

---

* We usually specify that the test mass m_0 must be small. That is, we do not want it to change the Earth's gravitational field. If, for instance, we used a test mass of the size of the Moon, its gravitational force on the Earth would cause tidal effects that would change the distribution of mass on the Earth and thus change the gravitational force at various locations. To prevent this from happening, we keep m_0 much smaller than the mass of the Earth.

TABLE 26-1 Some Electric Fields

Location	*Electric Field* (N/C)
At the surface of a uranium nucleus	3×10^{21}
In a hydrogen atom, at the electron's average radius	5×10^{11}
Electric breakdown occurs in air	3×10^{6}
At the charged drum of a photocopier	10^{5}
The electron beam accelerator in a TV set	10^{5}
Near a charged plastic comb	10^{3}
In the lower atmosphere	10^{2}
Inside the copper wire of household circuits	10^{-2}

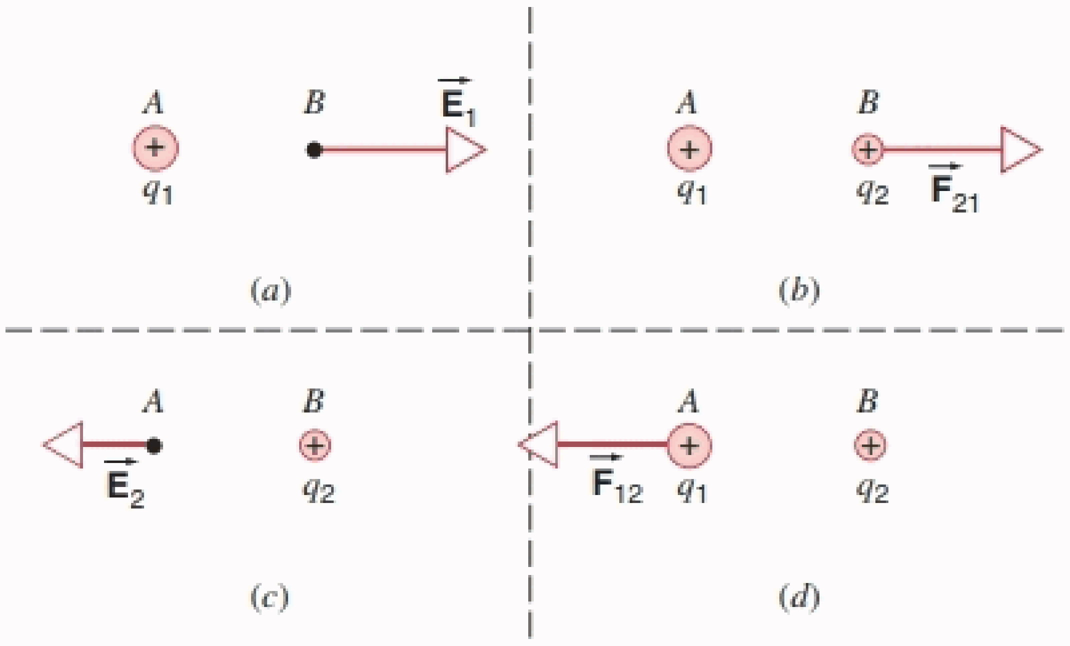


FIGURE 26-2. (*a*) q_1 at A sets up an electric field at B. (*b*) The electric field at B exerts a force on q_2. (*c*) q_2 at B sets up an electric field at A. (*d*) The electric field at A exerts a force on q_1. Note that $\vec{\mathbf{F}}_{12} = -\vec{\mathbf{F}}_{21}$.

Dimensionally, the electric field is the force per unit charge, and its SI unit is the newton/coulomb (N/C), although it is more often given, as we discuss in Chapter 28, in the equivalent unit of volt/meter (V/m). Note the similarity with the gravitational field, in which g (which is usually expressed in units of m/s^2) can also be expressed as the force per unit mass in units of newton/kilogram (N/kg). Both the gravitational and electric fields can be expressed as a force divided by a property (mass or charge) of the test body. Table 26-1 shows some electric fields that occur in a few situations.

As we did in Eq. 26-2 for the gravitational force on a body, we can use the electric field to calculate the force on any charged body. Once we have found the electric field at a point (using our test body, for instance), we can find the electric force exerted on any object of charge q at that location as

$$\vec{\mathbf{F}} = q\vec{\mathbf{E}}. \tag{26-4}$$

Here the electric field $\vec{\mathbf{E}}$ is caused by other charges that may be present, not by the charge q. Equation 26-4 is simply a way of specifying the force that these other charges exert on q.

We can now understand how the electric field acts as an intermediary in the interaction between two charges q_1 and q_2. As illustrated in Fig. 26-2*a*, charge q_1 located at A sets up an electric field at all surrounding points. Let $\vec{\mathbf{E}}_1$ be its value at location B. We can find the value of the field by placing our test charge at B and measuring the force exerted on it by q_1. If we place a different charge q_2 at B, it will experience an electric force $\vec{\mathbf{F}}_{21}$, which we can calculate using Eq. 26-4: $\vec{\mathbf{F}}_{21} = q_2\vec{\mathbf{E}}_1$ (Fig. 26-2*b*). The situation is completely symmetric: we could instead first use the test charge to determine the field $\vec{\mathbf{E}}_2$ at A due to q_2 (Fig. 26-2*c*). Then placing q_1 at A, we find the force exerted on q_1 by q_2: $\vec{\mathbf{F}}_{12} = q_1\vec{\mathbf{E}}_2$ (Fig. 26-2*d*). From Newton's third law the forces are equal and opposite ($\vec{\mathbf{F}}_{21} = -\vec{\mathbf{F}}_{12}$), even though the electric fields set up by the two charges are different.

To use Eq. 26-3 as an operational procedure for measuring the electric field, we must apply the same caution we did in using a test mass to measure the gravitational field: the test charge should be sufficiently small so that it does not disturb the distribution of charges whose electric field we are trying to measure. That is, we should more properly write Eq. 26-3 as

$$\vec{\mathbf{E}} = \lim_{q_0 \to 0} \frac{\vec{\mathbf{F}}}{q_0}, \tag{26-5}$$

even though we know from Chapter 25 that this limit in actuality cannot be taken to zero because the test charge can never be smaller than the elementary charge e. Of course, if we are *calculating* (rather than measuring) the electric field due to a specified collection of charges at fixed positions, neither the magnitude nor the sign of q_0 affects the result. As we show later in this chapter, electric fields of collections of charges can be calculated using Eq. 26-3 and Coulomb's law without direct reference to Eq. 26-5.

SAMPLE PROBLEM 26-1. An electron ($q = -e$) placed near a charged body experiences a force in the $+y$ direction of magnitude 3.60×10^{-8} N. (*a*) What is the electric field at that location? (*b*) What would be the force exerted by the same charged body on an alpha particle ($q = +2e$) placed at the location formerly occupied by the electron?

Solution (*a*) Using Eq. 26-4, we have

$$E_y = \frac{F_y}{q} = \frac{3.60 \times 10^{-8}\ \text{N}}{-1.60 \times 10^{-19}\ \text{C}} = -2.25 \times 10^{11}\ \text{N/C}.$$

The electric field is in the negative y direction.

(*b*) The force on the alpha particle follows from Eq. 26-4:

$$\begin{aligned} F_y = qE_y &= 2(+1.60 \times 10^{-19}\ \text{C})(-2.25 \times 10^{11}\ \text{N/C}) \\ &= -7.20 \times 10^{-8}\ \text{N}. \end{aligned}$$

The force is in the negative y direction, the same direction as the electric field but opposite to the direction of the force on the electron. In the same electric field, the force on the alpha particle is twice as large as the force on the electron, because the charge of the alpha particle is twice the magnitude of the electron's charge.

26-3 THE ELECTRIC FIELD OF POINT CHARGES

In this section we consider the electric field of point charges, first a single charge and then an assembly of individual charges. Later we generalize to continuous distributions of charge.

Let a positive test charge q_0 be placed a distance r from a point charge q. The magnitude of the force acting on q_0 is given by Coulomb's law,

$$F = \frac{1}{4\pi\epsilon_0}\frac{q_0|q|}{r^2}.$$

The magnitude of the electric field at the site of the test charge is, from Eq. 26-3,

$$E = \frac{F}{q_0} = \frac{1}{4\pi\epsilon_0}\frac{|q|}{r^2}. \qquad (26\text{-}6)$$

The direction of $\vec{\mathbf{E}}$ is the same as the direction of $\vec{\mathbf{F}}$, along a radial line from q, pointing outward if q is positive and inward if q is negative. Figure 26-3 shows the magnitude and direction of the electric field $\vec{\mathbf{E}}$ at various points near a positive point charge. How would this figure be drawn if the charge were negative?

To find $\vec{\mathbf{E}}$ for a group of N point charges, the procedure is as follows: (1) Calculate $\vec{\mathbf{E}}_n$ due to each charge n at the given point *as if it were the only charge present.* (2) Add these separately calculated fields vectorially to find the resultant field $\vec{\mathbf{E}}$ at the point. In equation form,

$$\begin{aligned}\vec{\mathbf{E}} &= \vec{\mathbf{E}}_1 + \vec{\mathbf{E}}_2 + \vec{\mathbf{E}}_3 + \cdots \\ &= \sum \vec{\mathbf{E}}_n \qquad (n = 1, 2, 3, \ldots, N). \qquad (26\text{-}7)\end{aligned}$$

The sum is a vector sum, taken over all the charges. Equation 26-7 (like Eq. 25-8) is an example of the application of the *principle of superposition,* which states, in this context, that at a given point the electric fields due to separate charge distributions simply add up (vectorially) or superimpose independently. This principle may fail when the magnitudes of the fields are extremely large, but it will be valid in all situations we discuss in this text.

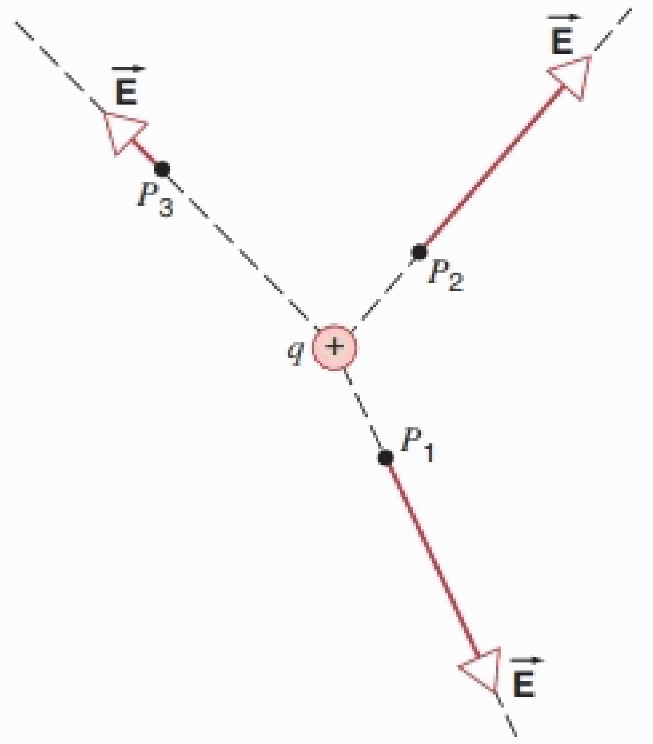


FIGURE 26-3. The electric field $\vec{\mathbf{E}}$ at various points near a positive point charge q. Note that the direction of $\vec{\mathbf{E}}$ is everywhere radially outward from q. The fields at P_1 and P_2, which are the same distance from q, are equal in magnitude. The field at P_3, which is twice as far from q as P_1 or P_2, has one-quarter the magnitude of the field at P_1 or P_2.

SAMPLE PROBLEM 26-2. In an ionized helium atom (a helium atom in which one of the two electrons has been removed), the electron and the nucleus are separated by a distance of 26.5 pm. What is the magnitude of the electric field due to the nucleus at the location of the electron?

Solution We use Eq. 26-6, with q (the charge of the nucleus) equal to $+2e$:

$$\begin{aligned}E &= \frac{1}{4\pi\epsilon_0}\frac{|q|}{r^2} = \frac{(8.99 \times 10^9\ \mathrm{N\cdot m^2/C^2})[2(1.60 \times 10^{-19}\ \mathrm{C})]}{(26.5 \times 10^{-12}\ \mathrm{m})^2} \\ &= 4.10 \times 10^{12}\ \mathrm{N/C}.\end{aligned}$$

This value is 8 times the electric field that acts on an electron in hydrogen (see Table 26-1). The increase comes about because (1) the nuclear charge in helium is twice that in hydrogen, and (2) the electron in ionized helium is closer to its nucleus (by a factor of two) than is the case for an electron in the hydrogen atom.

SAMPLE PROBLEM 26-3. Figure 26-4 shows a charge q_1 of $+1.5\ \mu\mathrm{C}$ and a charge q_2 of $+2.3\ \mu\mathrm{C}$. The first charge is at the origin of an x axis, and the second is at a position $x = L$, where $L = 13$ cm. At what point P along the x axis is the electric field zero?

Solution The point must lie between the charges because only in this region do the forces exerted by q_1 and q_2 on a test charge oppose each other. If $\vec{\mathbf{E}}_1$ is the electric field due to q_1 and $\vec{\mathbf{E}}_2$ is that due to q_2, the magnitudes of these vectors must be equal, or

$$E_1 = E_2.$$

From Eq. 26-6 we then have

$$\frac{1}{4\pi\epsilon_0}\frac{q_1}{x^2} = \frac{1}{4\pi\epsilon_0}\frac{q_2}{(L - x)^2},$$

where x is the coordinate of point P. Taking the square root of each side and solving for x, we obtain

$$x = \frac{L}{1 \pm \sqrt{q_2/q_1}},$$

where we have taken into account that the square root may have either a positive or a negative value. Substituting numerical values for L, q_1, and q_2, we obtain

$$x = 5.8\ \mathrm{cm} \quad \text{and} \quad x = -54.6\ \mathrm{cm}.$$

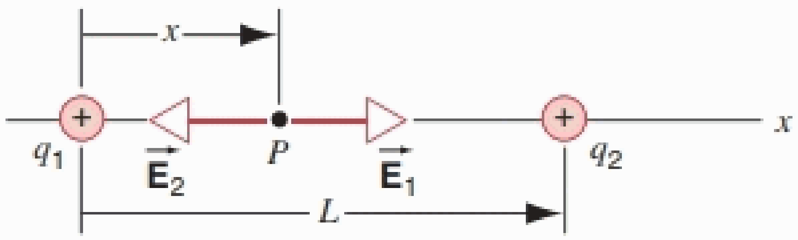


FIGURE 26-4. Sample Problem 26-3. At point P, the electric fields of the charges q_1 and q_2 are equal and opposite, so the net field at P is zero.

The first solution, which defines a point between the charges, is the solution we seek. The second solution defines a point to the left of the two charges. At this point it is true that $E_1 = E_2$, but the fields point in the same direction so their vector sum cannot be zero. Thus we are justified in discarding the second solution.

The Electric Dipole

Many objects found in nature can be successfully analyzed as isolated bodies having a net charge, as we have done so far in this chapter. Others show different kinds of behavior. One type of behavior is characteristic of an object that has no *net* charge but instead consists of equal positive and negative charges $+q$ and $-q$ separated by a fixed distance d. For example, an ionic molecule such as NaCl (in the high-temperature vapor state, not the familiar crystalline form) is electrically neutral but can be considered as a Na^+ ion joined to a Cl^- ion. For another example, a similar type of behavior of the water molecule is in part responsible for the large solubilities of many substances in water.

The configuration of two equal and opposite charges separated by a distance is called an *electric dipole*. In equations that describe electric dipoles, we find that the magnitude of the charge q on each of the components and their separation d often occur together as the product qd. It is convenient to define this quantity as the *electric dipole moment p:*

$$p = qd. \tag{26-8}$$

It turns out that this quantity behaves like a vector. We define the vector electric dipole moment to have a magnitude $p = qd$ and a direction pointing from the negative charge to the positive charge along the line joining the two charges. Figure 26-5*a* shows an electric dipole and its vector dipole moment. For example, in NaCl the magnitude of the charge q on each ion is e and the measured separation distance is 0.236 nm, so we expect the dipole moment of the molecule to be

$$\begin{aligned} p = ed &= (1.60 \times 10^{-19}\,\text{C})(0.236 \times 10^{-9}\,\text{m}) \\ &= 3.78 \times 10^{-29}\,\text{C}\cdot\text{m}. \end{aligned}$$

The measured value is $3.00 \times 10^{-29}\,\text{C}\cdot\text{m}$, indicating that the electron is not entirely removed from Na and attached to Cl. To a certain extent, the electron is shared between Na and Cl, resulting in a dipole moment somewhat smaller than expected.

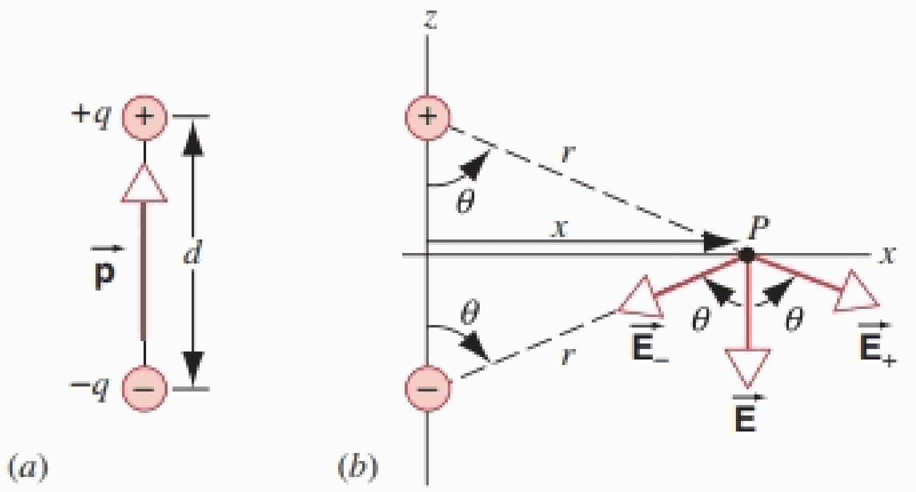


FIGURE 26-5. (*a*) Positive and negative charges of equal magnitude form an electric dipole. (*b*) The electric field $\vec{\mathbf{E}}$ at any point is the vector sum of the fields due to the individual charges. At point P on the x axis, the field has only a z component.

We now calculate the electric field $\vec{\mathbf{E}}$ of the dipole at point P a distance x along the perpendicular bisector of the dipole, as shown in Fig. 26-5*b*.

The positive and negative charges set up electric fields $\vec{\mathbf{E}}_+$ and $\vec{\mathbf{E}}_-$, respectively. The magnitudes of these two fields at P are equal, because P is equidistant from the positive and negative charges. Figure 26-5*b* also shows the directions of $\vec{\mathbf{E}}_+$ and $\vec{\mathbf{E}}_-$, determined by the directions of the force due to each charge alone that would act on a positive test charge at P. The total electric field at P is determined, according to Eq. 26-7, by the vector sum

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_+ + \vec{\mathbf{E}}_-.$$

From Eq. 26-6, the magnitudes of the fields from each charge are given by

$$E_+ = E_- = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{q}{x^2 + (d/2)^2}. \tag{26-9}$$

Because the fields $\vec{\mathbf{E}}_+$ and $\vec{\mathbf{E}}_-$ have equal magnitudes and lie at equal angles θ with respect to the z direction as shown, the x component of the total field is $E_+ \sin\theta - E_- \sin\theta = 0$. The total field $\vec{\mathbf{E}}$ therefore has only a z component, of magnitude

$$E = E_+ \cos\theta + E_- \cos\theta = 2E_+ \cos\theta. \tag{26-10}$$

From the figure we see that the cosine of the angle θ is determined according to

$$\cos\theta = \frac{d/2}{\sqrt{x^2 + (d/2)^2}}.$$

Substituting this result and Eq. 26-9 into Eq. 26-10, we obtain

$$E = (2)\frac{1}{4\pi\epsilon_0}\frac{q}{x^2 + (d/2)^2}\frac{d/2}{\sqrt{x^2 + (d/2)^2}}$$

or

$$E = \frac{1}{4\pi\epsilon_0}\frac{p}{[x^2 + (d/2)^2]^{3/2}}, \tag{26-11}$$

using Eq. 26-8 ($p = qd$) for the dipole moment.

Equation 26-11 gives the magnitude of the electric field at P due to the dipole. Note that the problem has cylindrical symmetry about the z axis; that is, we could have chosen the x axis to have any direction perpendicular to the dipole axis, and the field would be given by Eq. 26-11.

Often we observe the field of an electric dipole at points P whose distance x from the dipole is very large compared with the separation d. In this case we can simplify the dipole field somewhat by making use of the binomial expansion,

$$(1 + y)^n = 1 + ny + \frac{n(n-1)}{2!}y^2 + \cdots.$$

Let us rewrite Eq. 26-11 as

$$E = \frac{1}{4\pi\epsilon_0}\frac{p}{x^3}\frac{1}{[1+(d/2x)^2]^{3/2}}$$

$$= \frac{1}{4\pi\epsilon_0}\frac{p}{x^3}\left[1+\left(\frac{d}{2x}\right)^2\right]^{-3/2}$$

and apply the binomial expansion to the factor in brackets, which gives

$$E = \frac{1}{4\pi\epsilon_0}\frac{p}{x^3}\left[1+\left(-\frac{3}{2}\right)\left(\frac{d}{2x}\right)^2+\cdots\right].$$

For $x \gg d$ it is sufficient to keep only the first term in the brackets (the 1), and so we find an expression for the magnitude of the electric field due to a dipole at distant points in its median plane:

$$E = \frac{1}{4\pi\epsilon_0}\frac{p}{x^3}. \tag{26-12}$$

An expression of a similar form is obtained for the field along the dipole axis (the z axis of Fig. 26-5*b*); see Problem 1. A more general result for the field at any point in the xz plane can also be calculated; see Problem 2. In all cases, the field at distant points varies with the distance r from the dipole as $1/r^3$. This is a characteristic result for the electric dipole field. The field varies more rapidly with distance than the $1/r^2$ dependence characteristic of a point charge.

There are also more complicated charge distributions that give electric fields that vary as higher inverse powers of r. See Exercise 11 and Problem 4 for examples of the $1/r^4$ variation of the field of an electric *quadrupole.*

26-4 ELECTRIC FIELD OF CONTINUOUS CHARGE DISTRIBUTIONS

In Section 25-5 we discussed the force exerted on a point charge by various continuous charge distributions. We analyzed those continuous charge distributions by considering them to be collections of infinitesimal charge elements, which we treated as point charges, and then integrating over the distribution to find the force. We use a similar method here to calculate the electric field due to continuous charge distributions. In fact, as we shall see, we can use the results of Section 25-5 to obtain the electric field due to the distributions considered in that section.

We first discuss the general method for finding the electric field of a continuous charge distribution. We divide the charge distribution into infinitesimal elements dq, expressing the charge element dq as $\lambda\, ds$, $\sigma\, dA$, or $\rho\, dV$, depending on whether the charge is distributed over a line (λ = linear charge density or charge per unit length), surface (σ = surface charge density or charge per unit area), or volume (ρ = volume charge density or charge per unit volume). Choosing an arbitrary charge element, we write the magnitude of the contribution to the electric field at the observation point P as if dq were a point charge:

$$dE = \frac{1}{4\pi\epsilon_0}\frac{|dq|}{r^2}, \tag{26-13}$$

using Eq. 26-6. The direction of the vector $d\vec{\mathbf{E}}$ is determined by the sign of dq according to the direction of the force that dq would exert on a positive test charge at P. The total resultant field at P for the entire distribution is obtained by adding the contributions from all the charge elements of the object, taking into account the different directions that all the $d\vec{\mathbf{E}}$ might have:

$$\vec{\mathbf{E}} = \int d\vec{\mathbf{E}}. \tag{26-14}$$

In Cartesian coordinates, we can regard Eq. 26-14 as a shorthand representation of the three component equations:

$$E_x = \int dE_x, \qquad E_y = \int dE_y, \qquad E_z = \int dE_z. \tag{26-15}$$

As we discuss below, we can often simplify the calculation by arguing on the basis of symmetry that one or two of these integrals vanish or that two of them have identical values.

A Uniform Line of Charge

As an example of the application of Eqs. 26-13 to 26-15, we consider the electric field due to a line of charge (a thin charged rod, for example) of length L having a uniform positive linear charge density $\lambda = q/L$, where q is the total charge carried by the rod. Figure 26-6 shows the geometry for the calculation. We wish to find the field at point P a distance y from the rod along its perpendicular bisector (the positive y axis). The magnitude of the electric field $d\vec{\mathbf{E}}$ at point P due to the charge element dq is given by Eq. 26-13. We can conclude that $E_x = 0$, because none of the charge

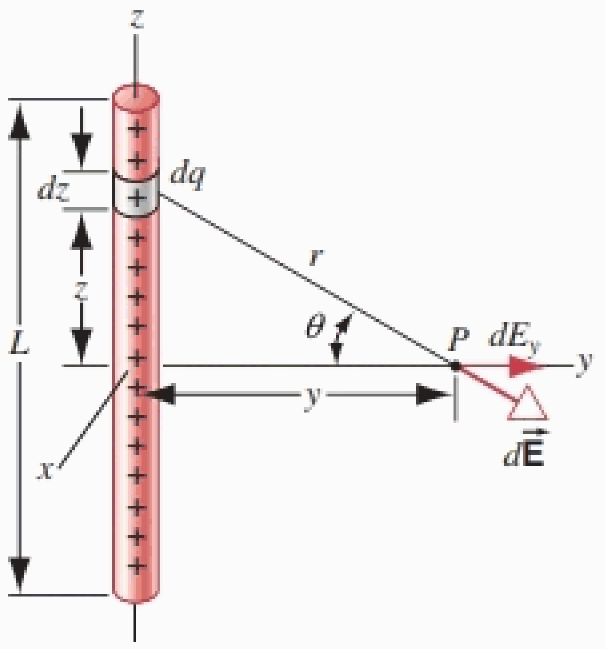


FIGURE 26-6. A uniformly charged rod. The electric field at point P is due to the total effect of all charge elements such as dq.

elements dq anywhere on the rod produces a $d\vec{\mathbf{E}}$ with an x component. We can also conclude from symmetry that $E_z = 0$, because for every dq at positive z there is a corresponding dq at negative z such that, when we add the $d\vec{\mathbf{E}}$ vectors from the two charge elements, the z components cancel. The only nonzero component of the electric field at P is E_y. We therefore have

$$dE_y = dE \cos\theta = \frac{1}{4\pi\epsilon_0}\frac{\lambda\, dz}{y^2 + z^2}\frac{y}{\sqrt{y^2 + z^2}},$$

where we have used Eq. 26-13 for dE with $dq = \lambda\ dz$, $\cos\theta = y/r$, and $r^2 = y^2 + z^2$. The total field at P is

$$E_y = \int dE_y = \int_{-L/2}^{+L/2} \frac{1}{4\pi\epsilon_0}\frac{\lambda y\, dz}{(y^2 + z^2)^{3/2}}.$$

Carrying out the integration over z, with y held constant, we obtain (see integral 18 in Appendix I)

$$E_y = \frac{1}{4\pi\epsilon_0}\frac{\lambda L}{y\sqrt{y^2 + L^2/4}}. \qquad (26\text{-}16)$$

This equation gives the electric field at point P on the positive y axis due to the line of charge. Note that we could have obtained this result directly from Eq. 25-15 for the force between the line of charge and the point charge q_0 by substituting λL for q and using Eq. 26-3, $E_y = F_y/q_0$.

As was the case with the electric dipole, this problem also has cylindrical symmetry about the z axis, and we could have chosen the y axis to point in any direction perpendicular to the axis of the rod and through its midpoint. Figure 26-7 shows a representation of the field in the xy plane due to a uniform positively charged rod.

As we did for the force calculated in Chapter 25, it is important to check our electric field calculations to verify that they have the correct limits. In the limit $y \to \infty$, Eq. 26-16 approaches the expression for the electric field of a point charge,

$$E_y = \frac{1}{4\pi\epsilon_0}\frac{q}{y^2},$$

where we have used $q = \lambda L$.

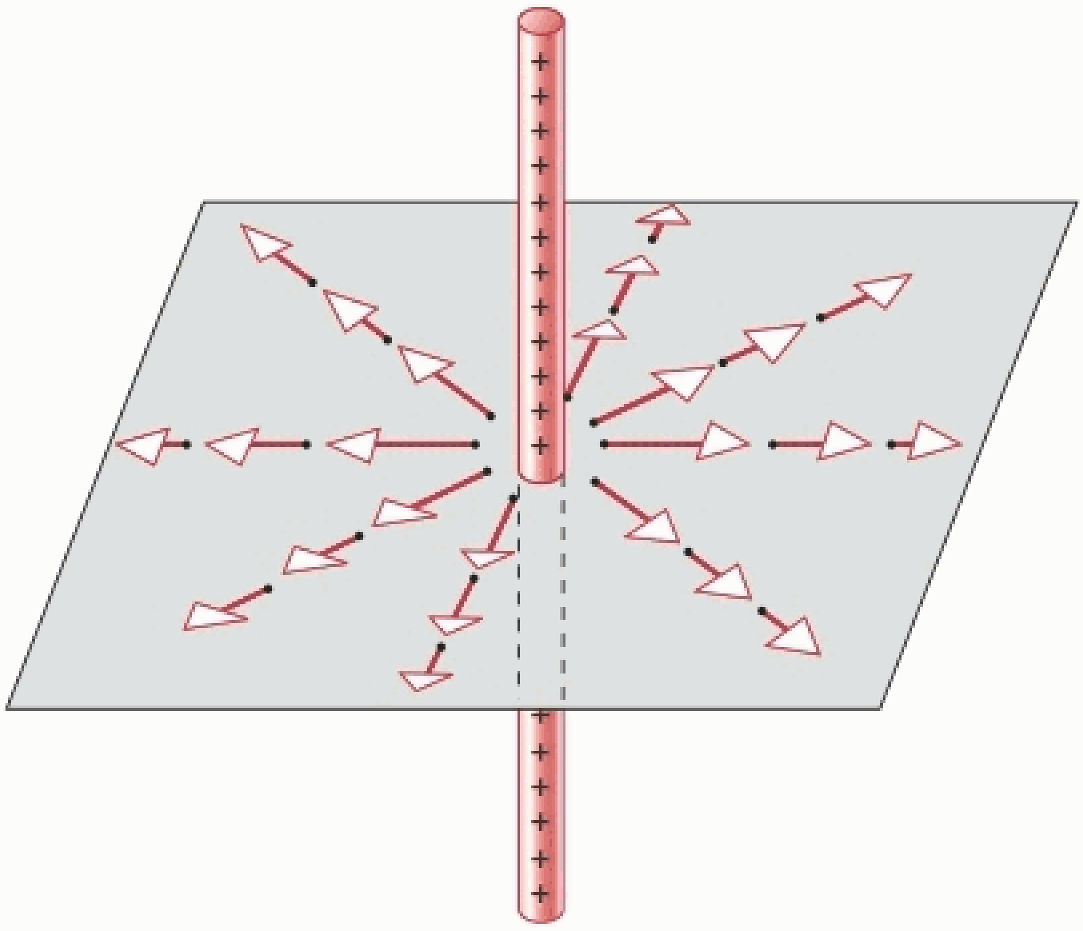


FIGURE 26-7. Electric field due to a positively charged rod. The field has cylindrical symmetry about the axis of the rod.

Often in situations involving line charges, our observation point is very close to the line, such that y is small in comparison with L. Taking the limit of Eq. 26-16 when $L \gg y$ with λ remaining constant, we have the electric field due to an infinitely long line of charge:

$$E_y = \frac{\lambda}{2\pi\epsilon_0 y}. \qquad (26\text{-}17)$$

The field is directed radially outward from the rod and depends inversely on the distance from the rod.

You may wonder about the usefulness of calculating the field due to an infinite line of charge when any real line of charge must have a finite length. However, for points close to the line and far from either end, Eq. 26-17 gives a very good and useful approximation to the electric field. The difference between the approximate result, Eq. 26-17, and the exact result, Eq. 26-16, is often negligible. The approximate result in this case may give more physical insight, because the variation of E with distance from the rod is more immediately apparent.

A Uniform Ring or Disk of Charge

To discuss the electric field due to a ring or disk of radius R carrying a uniform charge density, it is not necessary to do the complete calculation starting with Eq. 26-13. We have already calculated the force exerted on a point charge q_0 by a ring of charge or a disk of charge. The force exerted by a ring of charge on a point charge q_0 on the axis of the ring was given by Eq. 25-16. Using Eq. 26-3, $E_z = F_z/q_0$, we can find the electric field at a point on the positive z axis due to a ring of charge directly from Eq. 25-16:

$$E_z = \frac{\lambda}{2\epsilon_0}\frac{Rz}{(z^2 + R^2)^{3/2}}, \qquad (26\text{-}18)$$

where we have used $q = \lambda(2\pi R)$. The electric field is directed along the axis of the ring (the z axis) and away from the ring. Equation 26-18 is valid for positive as well as negative z. If the ring is negatively charged, the field points along the axis in the opposite direction (toward the ring).

In a similar fashion, we can find the electric field due to a disk of charge from Eq. 25-17:

$$E_z = \frac{\sigma}{2\epsilon_0}\left(1 - \frac{z}{\sqrt{z^2 + R^2}}\right). \qquad (26\text{-}19)$$

Here we have expressed the electric field in terms of the surface charge density of the disk using $q = \sigma A = \sigma(\pi R^2)$. This expression gives the field at a point on the positive z

axis a distance z from the disk. The field points away from the disk if the disk is positively charged. Equation 26-19 is valid only for $z > 0$. How should it be modified if P is located on the negative z axis?

An Infinite Sheet of Charge

Let us now consider the limiting case of Eq. 26-19 as $R \rightarrow \infty$, so that the charged disk becomes an infinite sheet of charge. We assume that, as R increases, we add charge to the disk so that the surface charge density σ remains constant. Under these conditions, we can approximate Eq. 26-19 as

$$E_z = \frac{\sigma}{2\epsilon_0}. \qquad (26\text{-}20)$$

This turns out to be a very useful result, which is approximately valid for a disk of uniform charge density when we are close to the disk and far from any of its edges. In fact, if we are far from its edges, we cannot tell whether the charge distribution is spread over a circular area or over one that is square, rectangular, or irregularly shaped. As we will derive in the next chapter, this result is valid for any large uniformly charged sheet, no matter what its shape. The field has a uniform magnitude and (for a positively charged sheet) is directed away from the sheet of charge.

A Uniform Spherical Shell of Charge

In Section 25-5, we established two properties of a uniformly charged spherical shell: it exerts no force on a test charge in its interior, and at exterior points the force that it exerts on a test charge is the same as if all the charge of the shell were concentrated in a point at its center. We can use these properties to deduce the electric field due to a thin uniformly charged shell. Let the shell have radius R and charge q, which we assume for now to be positive. We then have the following results for the electric field at various distances from the center of the shell:

$$E = 0 \qquad (r < R) \qquad (26\text{-}21a)$$

$$E_r = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} \qquad (r \geq R). \qquad (26\text{-}21b)$$

The subscript r on the electric field reminds us that the field points in the radial direction. These results follow directly from the force on a test charge at the different locations. Inside the shell, the electric field is zero. At exterior points, the electric field is radial and identical to that of a point charge, so it would look just like the field displayed in Fig. 26-3.

We can use the properties of shells of charge to deduce the electric field due to a spherically symmetric charge distribution in a sphere of radius R. For simplicity we assume the charge to be distributed uniformly throughout the sphere, so that its volume charge density is a constant. If Q is the total charge on the sphere, then the volume charge density is

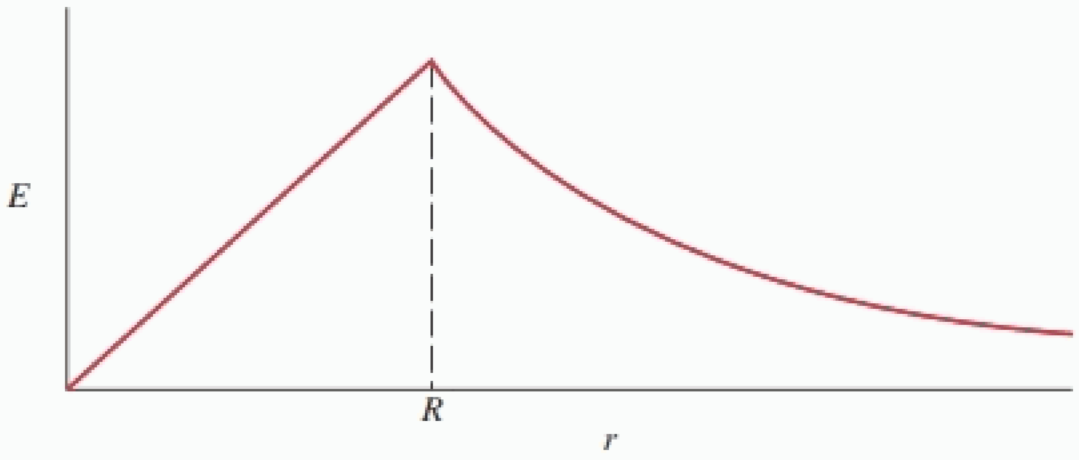


FIGURE 26-8. The magnitude of the electric field due to a uniformly charged sphere of radius R.

$$\rho = \frac{Q}{\frac{4}{3}\pi R^3}. \qquad (26\text{-}22)$$

Imagine the sphere to be divided into many thin shells of radius r and thickness dr. If we place a test charge at a distance r from the origin and inside the shell ($r < R$), the electric field at the location of the test charge is due only to the shells at smaller radii; we know from Eq. 26-21a that $E = 0$ for all the shells of larger radii. Furthermore, we know from Eq. 26-21b that the field due to all the shells of smaller radii is the same as that of a point charge at the origin. The magnitude of that point charge is the same as the total charge of all the shells with radii smaller than r, or equivalently the total charge q inside the sphere of radius r, which is given by the volume charge density times the volume of the sphere of radius r:

$$q = \rho(\tfrac{4}{3}\pi r^3) = Q\frac{r^3}{R^3}, \qquad (26\text{-}23)$$

using the charge density from Eq. 26-22.

Equation 26-21b then gives the radial component of the electric field at this location inside the sphere:

$$E_r = \frac{1}{4\pi\epsilon_0}\frac{Qr}{R^3} \qquad (r < R). \qquad (26\text{-}24)$$

If, as we assumed, Q is positive, the field points radially outward; if Q is negative, the field is radially inward. For $r > R$, the field is identical to that of a point charge Q located at the origin (Eq. 26-6). Note that the field increases linearly with r for $r < R$ and decreases like $1/r^2$ for $r > R$. Figure 26-8 shows the magnitude of the electric field as a function of r.

SAMPLE PROBLEM 26-4. A plastic rod, whose length L is 220 cm and whose radius R is 3.6 mm, carries a negative charge q of magnitude 3.8×10^{-7} C, spread uniformly over its surface. What is the electric field near the midpoint of the rod, at a point on its surface?

Solution Although the rod is not infinitely long, for a point on its surface and near its midpoint it is effectively very long, so that we are justified in using Eq. 26-17. The linear charge density for the rod is

$$\lambda = \frac{q}{L} = \frac{-3.8 \times 10^{-7}\,\mathrm{C}}{2.2\,\mathrm{m}} = -1.73 \times 10^{-7}\,\mathrm{C/m}.$$

From Eq. 26-17 we then have, for $y = 0.0036$ m,

$$\begin{aligned} E_y &= \frac{\lambda}{2\pi\epsilon_0 y} \\ &= \frac{-1.73 \times 10^{-7}\,\mathrm{C/m}}{(2\pi)(8.85 \times 10^{-12}\,\mathrm{C^2/N\cdot m^2})(0.0036\,\mathrm{m})} \\ &= -8.6 \times 10^5\,\mathrm{N/C}. \end{aligned}$$

The negative sign tells us that, because the rod is negatively charged, the direction of the electric field is radially inward, toward the axis of the rod. Sparking occurs in dry air at atmospheric pressure at an electric field strength of about 3×10^6 N/C. The field strength we calculated is lower than this by a factor of about 3.4, so that sparking should not occur.

26-5 ELECTRIC FIELD LINES

The concept of the electric field was introduced in the early 19th century by Michael Faraday. Faraday did not develop the mathematical representation of the electric field; instead, he developed a graphical representation, in which he imagined the space around an electric charge to be filled with *lines of force.* Today we no longer attach the same reality to the lines of force that Faraday did, but we retain them as a convenient way to visualize the electric field. We refer to these lines as *electric field lines.*

Figure 26-9*a* shows the electric field lines representing a uniform field. Note that the lines are parallel and equally spaced. Figure 26-9*b* shows lines representing a nonuniform field. By convention, we draw the field lines with the following property:

> *The tangent to the electric field line passing through any point in space gives the direction of the electric field at that point.*

In Fig. 26-9*a*, for example, the direction of the electric field at point P is vertically upward, tangent to the field lines. Because the field is uniform, the electric field has this direction at every point in this region of space. In Fig. 26-9*b*, which shows a nonuniform field, the electric field has different directions at points P_1 and P_2, in each case tangent to the electric field line passing through that point.

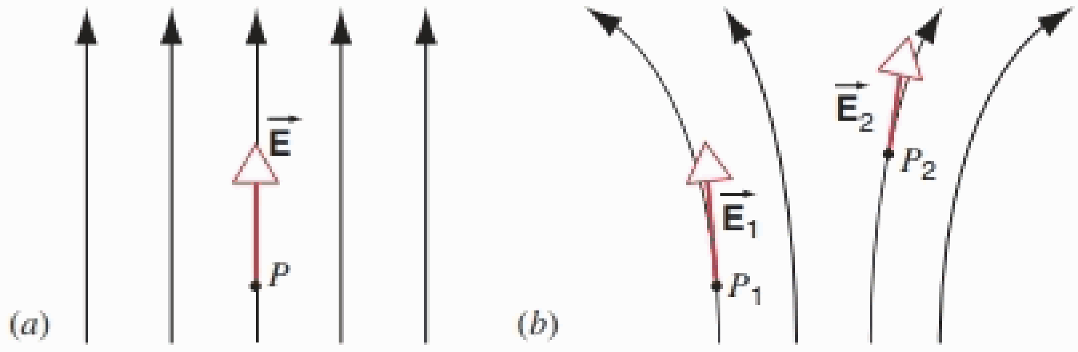


FIGURE 26-9. (*a*) Electric field lines for a uniform field. (*b*) Electric field lines for a nonuniform field.

For the electric field lines to have this property, they must also be drawn so that

> *The electric field lines start on positive charges and end on negative charges.*

For example, Fig. 26-10 represents the field lines for an isolated positive point charge (or a small sphere of positive charge). The lines point radially outward, so that at any point P the field is radial. The field lines begin on the positive charge and extend to infinity, since there are no negative charges in this region. If the charge were negative, the field lines would point in the opposite direction (radially inward).

One final property of electric field lines is that

> *The magnitude of the electric field at any point is proportional to the number of field lines per unit cross-sectional area perpendicular to the lines.*

In other words, the more densely packed the field lines are near any point, the stronger is the field at that point. Figure 26-9*b*, for example, suggests that the magnitude of the field is larger at the bottom of the drawing (near point P_1) than it is at the top of the drawing (near point P_2). In Fig. 26-9*a*, on the other hand, the spacing of the field lines is the same at all points, suggesting that the field has the same magnitude everywhere. For a point charge (Fig. 26-10), the field lines are close together near the charge and further apart away from the charge, which indicates that the field grows weaker as the distance from the charge increases.

The uniform field near a large sheet of positive charge is shown in Fig. 26-11. The direction of the field is perpendicular to the sheet. Near the edges of the sheet, the field becomes nonuniform and is no longer directed perpendicular to the sheet, but as long as we stay close to the center of the sheet and far away from any edge, the field is very nearly uniform. Once again, the field lines extend to infinity.

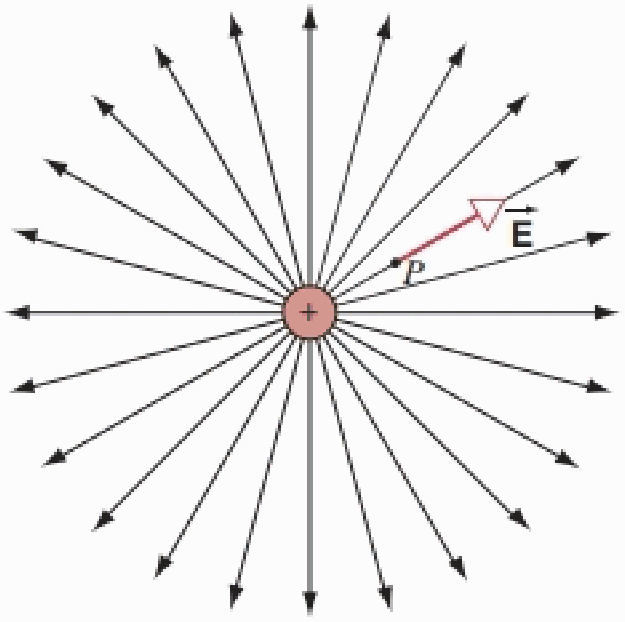


FIGURE 26-10. Electric field lines surrounding an isolated positive point charge or positively charged uniform sphere. The field at an arbitrary point P is shown.

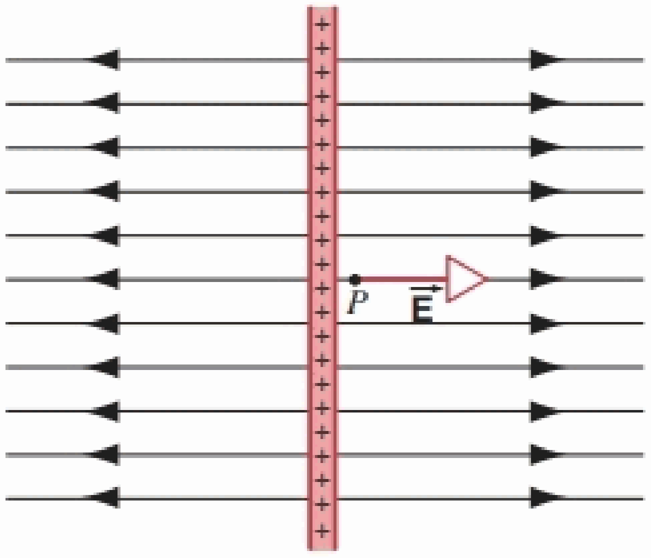


FIGURE 26-11. Electric field lines near a thin uniform sheet of charge. Here we are looking at the edge of the sheet, which is oriented perpendicular to the page.

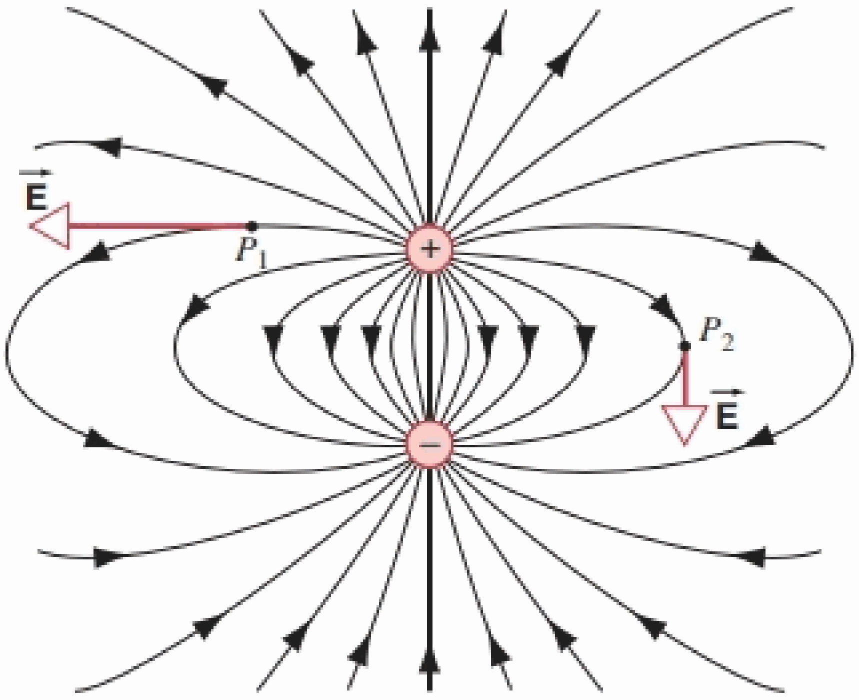


FIGURE 26-12. Electric field lines for an electric dipole.

Figure 26-12 shows the field near a dipole (which illustrates how the field lines begin on positive charges and end on negative charges), and Fig. 26-13 shows the field near two equal positive charges. Note the differences between the two patterns. In the region directly between the charges, the density of field lines is larger in Fig. 26-12 than in Fig. 26-13, suggesting that the dipole gives a larger field in that region than the two positive charges. Because the electric field is zero halfway between the charges in Fig. 26-13, no electric field lines can be drawn through that point. As we travel far from the charges in Fig. 26-13, the pattern begins

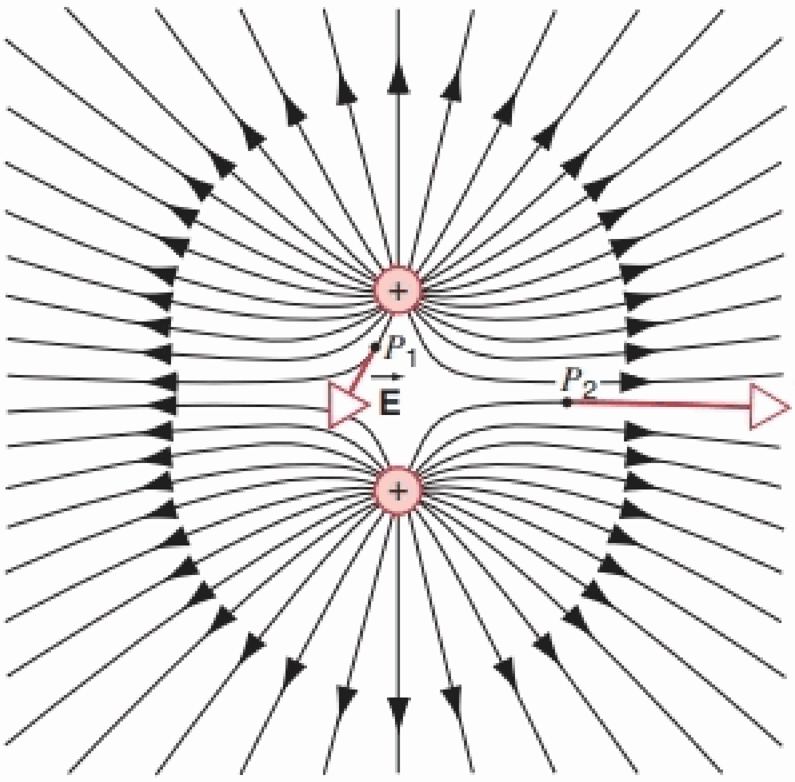
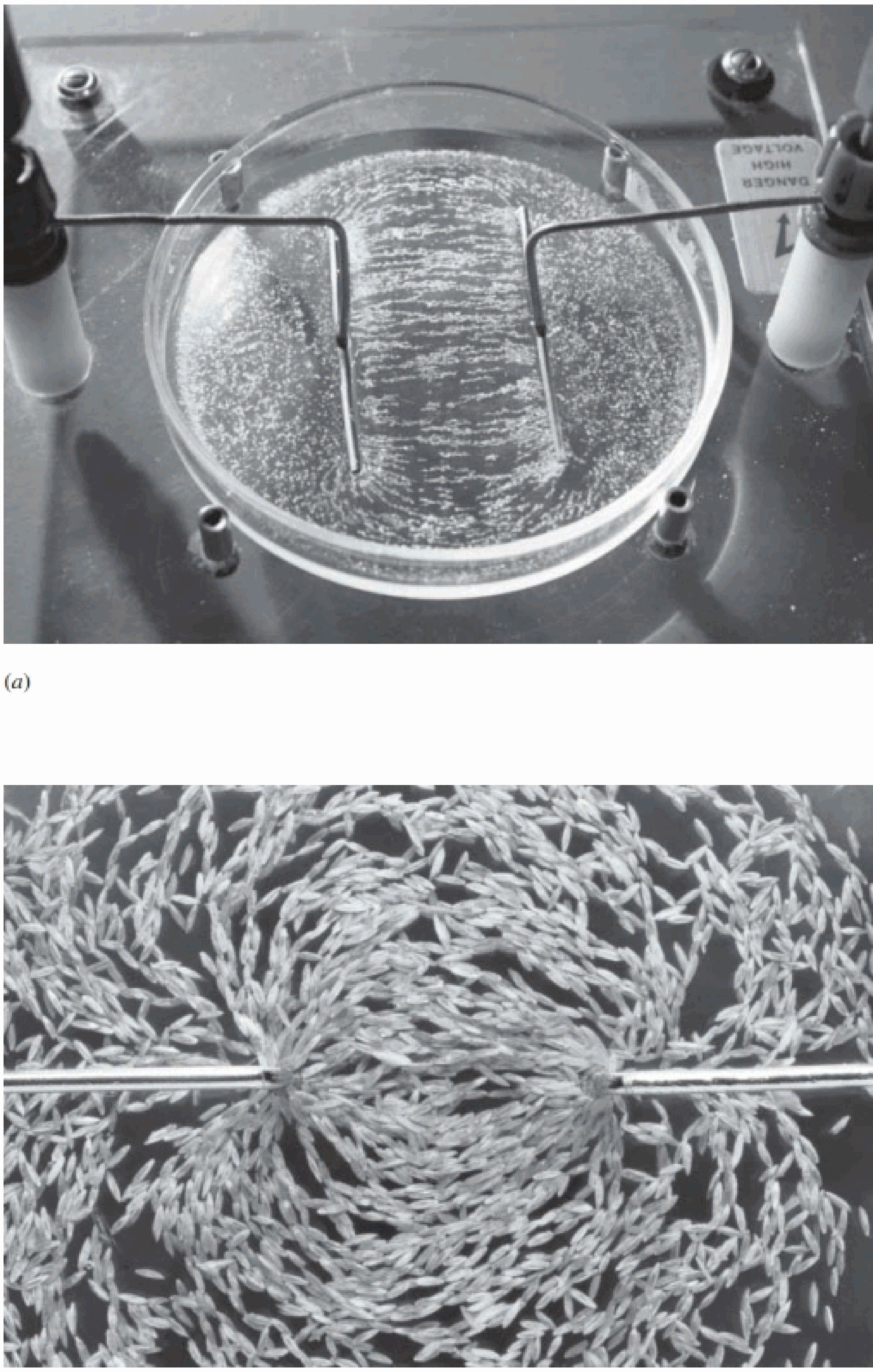


FIGURE 26-13. Electric field lines for two equal positive charges.


(*a*)

(*b*)

FIGURE 26-14. Photographs of the patterns of electric field lines around (*a*) a charged plate (which produces parallel field lines) and (*b*) two rods with equal and opposite charges (similar to the electric dipole of Fig. 26-12). The patterns were made visible by suspending grass seed in an insulating liquid.

to resemble that of a single charge (as in Fig. 26-10). At external points in the median plane (the plane perpendicular to the page and midway between the charges), the field is small for the dipole and directed downward, while for the equal charges the field is larger and directed radially outward, as indicated at the points P_2 in Figs. 26-12 and 26-13.

These drawings can be very useful in helping us to visualize the pattern of electric field lines. However, keep in mind that they represent only a two-dimensional "slice" through what is in reality a three-dimensional pattern. The relative spacings of the field lines in two dimensions do not strictly correspond with the spacings of the three-dimensional pattern, and the spacings of field lines in our two-dimensional drawings have no direct mathematical relationship to the magnitude of the field, other than to suggest regions where the field may be uniform or may be increasing or decreasing in magnitude.

The pattern of electric field lines can be made visible by applying an electric field to a suspension of tiny objects in an insulating fluid. Figure 26-14 shows photographs of patterns that resemble the drawings of electric field lines for a charged sheet and an electric dipole.

26-6 A POINT CHARGE IN AN ELECTRIC FIELD

In preceding sections we have considered the first part of the charge ⇄ field ⇄ charge interaction: Given a collection of charges, what is the resulting electric field? In this section and the next, we consider the second part: What happens when we put a charged particle in a known electric field?

From Eq. 26-4, we know that a particle of charge q in an electric field $\vec{\mathbf{E}}$ experiences a force $\vec{\mathbf{F}}$ given by

$$\vec{\mathbf{F}} = q\vec{\mathbf{E}}.$$

To study the motion of the particle in the electric field, all we need do is use Newton's second law, $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$, where the resultant force on the particle includes the electric force and any other forces that may act.

As we did in our original study of Newton's laws, we can achieve a simplification if we consider the case in which the force is constant. We therefore begin by considering cases in which the electric field and the corresponding electric force are uniform (that is, they do not vary with location) and constant (they do not vary with time). Such a situation can be achieved in practice in the region near a large uniform sheet of charge, as we discussed in Section 26-4. For even greater uniformity, we can use a pair of closely spaced sheets of opposite charge, obtained by connecting the terminals of a battery to a pair of parallel metal plates. In the following sample problems, we assume that the field exists only in the region between the plates and drops suddenly to zero when the particle leaves that region. In reality the field decreases rapidly over a distance that is of the order of the spacing between the plates; when this distance is small, we do not make too large an error in calculating the motion of the particle if we ignore the edge effect.

SAMPLE PROBLEM 26-5. A charged drop of oil of radius $R = 2.76\ \mu\text{m}$ and density $\rho = 918\ \text{kg/m}^3$ is maintained in equilibrium under the combined influence of its weight and a downward uniform electric field of magnitude $E = 1.65 \times 10^6$ N/C (Fig. 26-15). (*a*) Calculate the magnitude and sign of the charge on the drop. Express the result in terms of the elementary charge e. (*b*) The drop is exposed to a radioactive source that emits electrons. Two electrons strike the drop and are captured by it, changing its charge by two units. If the electric field remains at its constant value, calculate the resulting acceleration of the drop.

Solution (*a*) To keep the drop in equilibrium, its weight mg must be balanced by an equal electric force of magnitude qE acting upward. Because the electric field is given as being in the downward direction, the charge q on the drop must be negative for the electric force to point in a direction opposite the field. The equilibrium condition is

$$\sum\vec{\mathbf{F}} = m\vec{\mathbf{g}} + q\vec{\mathbf{E}} = 0.$$

Taking y components, we obtain

$$-mg + q(-E) = 0$$

or, solving for the unknown q,

$$\begin{aligned} q &= -\frac{mg}{E} = -\frac{\frac{4}{3}\pi R^3\rho g}{E} \\ &= -\frac{\frac{4}{3}\pi(2.76\times10^{-6}\ \text{m})^3(918\ \text{kg/m}^3)(9.80\ \text{m/s}^2)}{1.65\times10^6\ \text{N/C}} \\ &= -4.80\times10^{-19}\ \text{C}. \end{aligned}$$

If we write q in terms of the electronic charge $-e$ as $q = n(-e)$, where n is the number of electronic charges on the drop, then

$$n = \frac{q}{-e} = \frac{-4.80\times10^{-19}\ \text{C}}{-1.60\times10^{-19}\ \text{C}} = 3.$$

(*b*) If we add two additional electrons to the drop, its charge will become

$$q' = (n+2)(-e) = 5(-1.60\times10^{-19}\ \text{C}) = -8.00\times10^{-19}\ \text{C}.$$

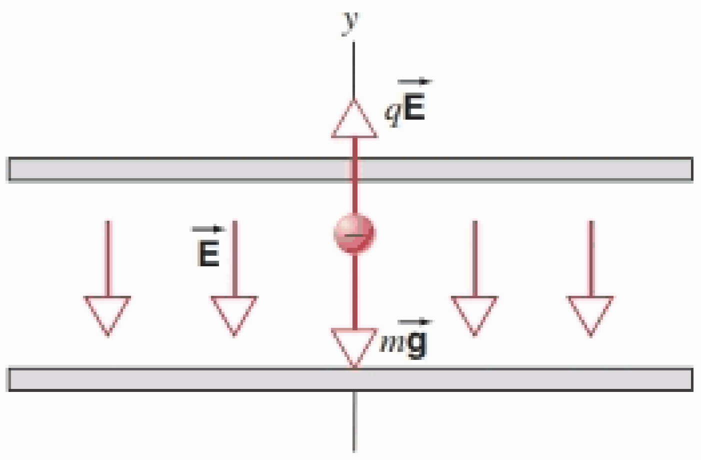


FIGURE 26-15. Sample Problem 26-5. A negatively charged drop is placed in a uniform electric field $\vec{\mathbf{E}}$. The drop moves under the combined influence of its weight $m\vec{\mathbf{g}}$ and the electric force $q\vec{\mathbf{E}}$.

Newton's second law can be written

$$\sum \vec{\mathbf{F}} = m\vec{\mathbf{g}} + q'\vec{\mathbf{E}} = m\vec{\mathbf{a}}$$

and, taking y components, we obtain

$$-mg + q'(-E) = ma.$$

We can now solve for the acceleration:

$$\begin{aligned} a &= -g - \frac{q'E}{m} \\ &= -9.80\ \text{m/s}^2 - \frac{(-8.00 \times 10^{-19}\ \text{C})(1.65 \times 10^6\ \text{N/C})}{\frac{4}{3}\pi(2.76 \times 10^{-6}\ \text{m})^3(918\ \text{kg/m}^3)} \\ &= -9.80\ \text{m/s}^2 + 16.3\ \text{m/s}^2 = +6.5\ \text{m/s}^2. \end{aligned}$$

The drop accelerates in the positive y direction.

In this calculation, we have ignored the viscous drag force, which is usually quite important in this situation. We have, in effect, found the acceleration of the drop at the instant it acquired the extra two electrons. The drag force, which depends on the velocity of the drop, is initially zero if the drop starts from rest, but it increases as the drop begins to move, and so the acceleration of the drop will decrease in magnitude.

This experimental configuration forms the basis of the Millikan oil-drop experiment, which was used to measure the magnitude of the electronic charge. The experiment is discussed later in this section.

SAMPLE PROBLEM 26-6. Figure 26-16 shows the deflecting electrode system of an ink-jet printer. An ink drop whose mass m is 1.3×10^{-10} kg carries a charge q of -1.5×10^{-13} C and enters the deflecting plate system with a speed $v = 18$ m/s. The length L of these plates is 1.6 cm, and the magnitude of the electric field E between the plates is 1.4×10^6 N/C. What is the vertical deflection of the drop at the far edge of the plates? Ignore the varying electric field at the edges of the plates.

Solution Let t be the time of passage of the drop through the deflecting system. The vertical and the horizontal displacements are given by

$$y = \tfrac{1}{2}at^2 \quad \text{and} \quad L = vt,$$

respectively, in which a is the vertical acceleration of the drop.

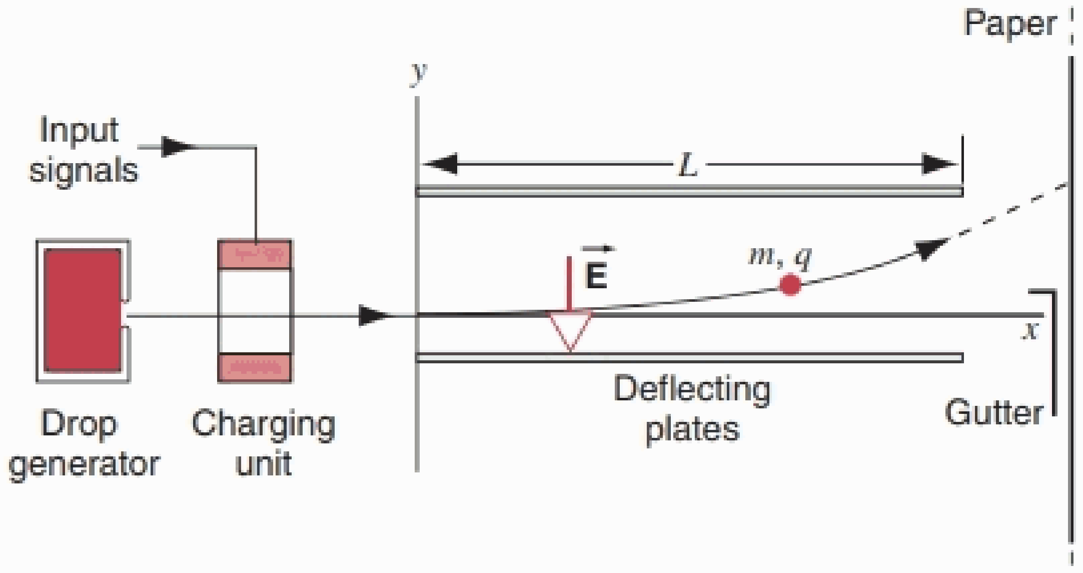


FIGURE 26-16. Sample Problem 26-6. The essential features of an ink-jet printer. An input signal from a computer controls the charge given to the drop and thus the position at which the drop strikes the paper. A transverse force from the electric field $\vec{\mathbf{E}}$ is responsible for deflecting the drop. The drop moves in a parabolic path while it is between the plates, and it moves along a straight line (shown dashed) after it leaves the plates.

As in the previous sample problem, we can write the y component of Newton's second law as $-mg + q(-E) = ma$. As you can easily verify, the electric force acting on the drop, $-qE$, is much greater in this case than the gravitational force mg so that the acceleration of the drop can be taken to be $-qE/m$. Eliminating t between the two displacement equations and substituting this value for a leads to

$$\begin{aligned} y &= \frac{-qEL^2}{2mv^2} \\ &= \frac{-(-1.5 \times 10^{-13}\ \text{C})(1.4 \times 10^6\ \text{N/C})(1.6 \times 10^{-2}\ \text{m})^2}{(2)(1.3 \times 10^{-10}\ \text{kg})(18\ \text{m/s})^2} \\ &= 6.4 \times 10^{-4}\ \text{m} = 0.64\ \text{mm}. \end{aligned}$$

The deflection at the paper will be larger than this because the ink drop follows a straight-line path to the paper after leaving the deflecting region, as shown by the dashed line in Fig. 26-16. To aim the ink drops so that they form the characters well, it is necessary to control the charge q on the drops—to which the deflection is proportional—to within a few percent. In our treatment, we have again neglected the viscous drag forces that act on the drop; they are substantial at these high drop speeds.

Measuring the Elementary Charge

We know today that electric charge is quantized; that is, it appears only in integral multiples of the elementary charge e, whose currently accepted value is $1.602176462 \times 10^{-19}$ C with an experimental uncertainty that shows up only in the last two digits. This modern value, like that of nearly all fundamental constants of physics, has been obtained from a variety of interlocking and increasingly precise experiments.

How did we first learn that charge was quantized, and how was the value of e first measured? The earliest definitive answers to these questions were obtained from experiments done by the American physicist Robert A. Millikan* (1868–1953). For this and related work Millikan received the 1923 Nobel Prize in physics.

Figure 26-17 shows Millikan's apparatus. An atomizer introduces oil droplets into chamber A. Some of the drops can become charged (positive or negative) in the process. We consider a drop of charge q (assumed negative); this drop drifts into chamber C through a small hole in plate P_1.

If there is no electric field in chamber C, two forces act on the drop, its weight mg and an upwardly directed viscous drag force, whose magnitude is proportional to the speed of the falling drop. The drop quickly comes to a

---

*For details of Millikan's experiments, see Henry A. Boorse and Lloyd Motz (eds.), *The World of the Atom* (Basic Books, 1966), Chapter 40. For the point of view of two physicists who knew Millikan as graduate students, see "Robert A. Millikan, Physics Teacher," by Alfred Romer, *The Physics Teacher,* February 1978, p. 78, and "My Work with Millikan on the Oil-Drop Experiment," by Harvey Fletcher, *Physics Today,* June 1982, p. 43.

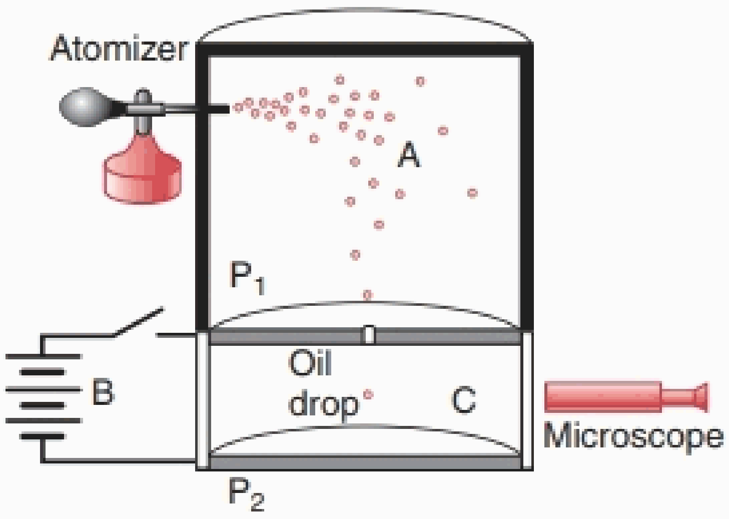


FIGURE 26-17. The Millikan oil-drop apparatus for measuring the elementary charge e. The motion of a drop is observed in chamber C, where the drop is acted on by gravity, the electric field set up by the battery B, and, if the drop is moving, a viscous drag force.

constant terminal speed v at which these two forces are just balanced.

A downward electric field $\vec{\mathbf{E}}$ is now set up in the chamber by connecting battery B between plates P$_1$ and P$_2$. A third force, $q\vec{\mathbf{E}}$, now acts on the drop. If q is negative, this force points upward, and—we assume—the drop now drifts upward, at a new terminal speed v'. In each case, the drag force points in the direction opposite to that in which the drop is moving and has a magnitude proportional to the speed of the drop. The charge q on the drop can be found from measurements of v and v'.

Millikan found that the values of q were all consistent with the relation

$$q = ne \qquad n = 0, \pm 1, \pm 2, \pm 3, \ldots .$$

That is, the charges on the drops occurred only in integer multiples of a certain fundamental quantity, the elementary charge e, which Millikan deduced to have a value of 1.64×10^{-19} C, completely consistent with the currently accepted value. Millikan's experiment provides convincing proof that charge is quantized.

Motion in Nonuniform Electric Fields (Optional)

So far we have considered only uniform fields, in which the electric field is constant in magnitude and direction over the region in which the particle moves. Often, however, we must deal with fields that are not uniform. For example, we consider a ring of positive charge, as shown in Fig. 26-18. The electric field on the axis of the ring is given by Eq. 26-18. Suppose we project a charged particle with initial speed v_0 along the z axis toward the ring from a large distance. As the particle moves along the axis, the electric field (and therefore the electric force on the particle) increases.

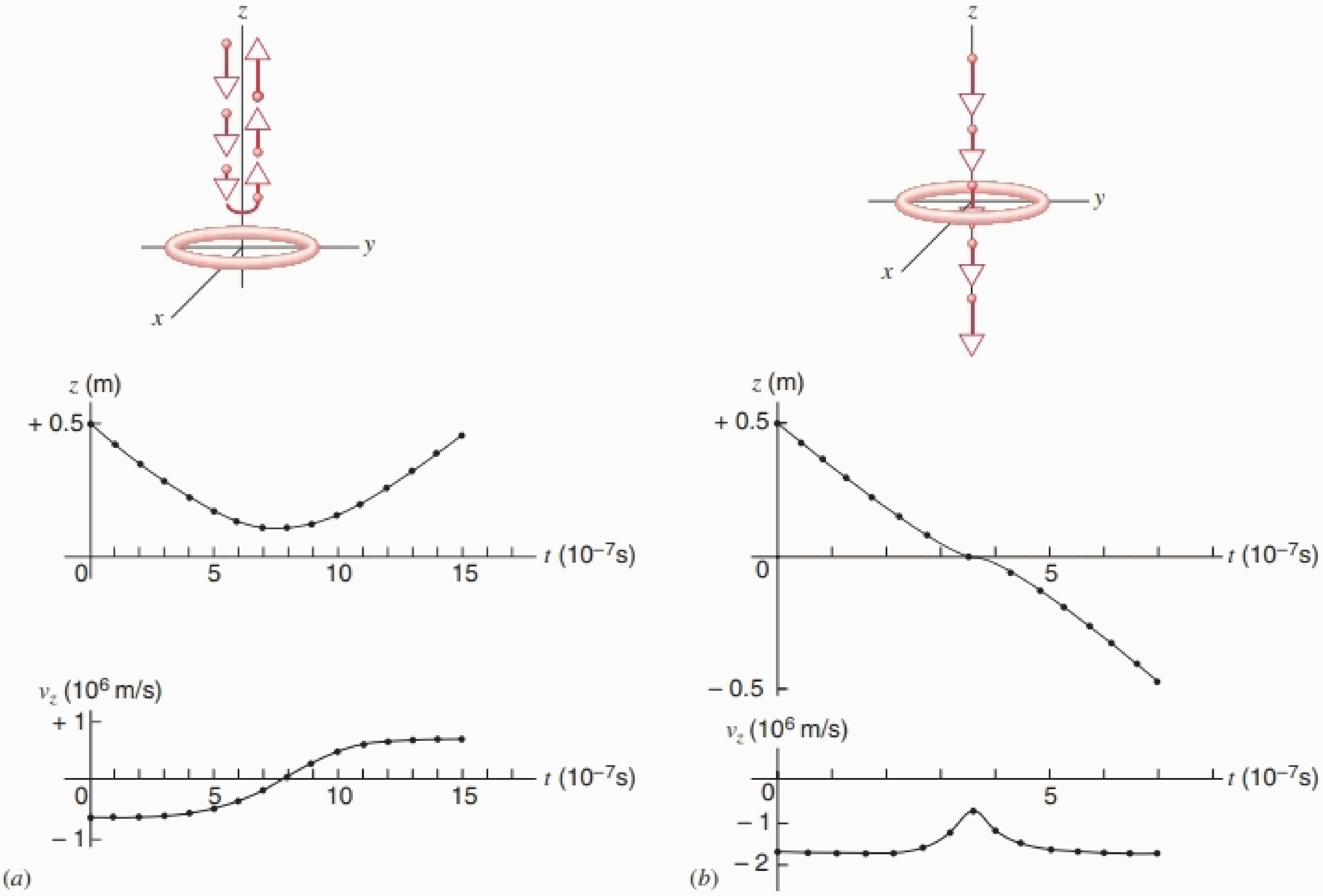


FIGURE 26-18. (*a*) The motion of a proton projected along the axis of a uniform positively charged ring. The position and velocity are shown. The proton comes instantaneously to rest at a time of about 8×10^{-7} s and reverses its motion. The points are the results of a numerical calculation; the curves are drawn through the points. (*b*) If the initial velocity of the proton is increased sufficiently, it can pass through the ring; its speed is a minimum as it passes through the center of the ring.

Neglecting gravity and considering only the electric force on the particle, how can we analyze its subsequent motion?

In such cases, we must use analytical methods for position-dependent forces similar to those discussed in Section 5-5 for time-dependent forces. An equivalent method is to follow the procedure given in Section 12-5 because, as we discuss in Chapter 28, the electrostatic force is a conservative force. Alternatively, we can use numerical techniques to find the solution by dividing the motion into infinitesimally small intervals over which we can take the acceleration to be nearly constant; an approximate solution can be obtained with a computer.

For this calculation, we use a ring of radius $R = 3$ cm and linear charge density $\lambda = +2 \times 10^{-7}$ C/m. A proton ($q = +1.6 \times 10^{-19}$ C, $m = 1.67 \times 10^{-27}$ kg) is projected along the axis of the loop from an initial position at $z = +0.5$ m with initial velocity $v_{z0} = -7 \times 10^5$ m/s. (The negative initial velocity means that the proton is moving downward toward the loop, which lies in the xy plane.) The positively charged loop exerts a repulsive force on the positively charged proton, decreasing its speed. In Fig. 26-18*a* we plot the resulting motion in the case that the proton does not have enough initial kinetic energy to reach the plane of the loop. The proton comes instantaneously to rest at a point just above the plane of the loop and then reverses its motion as the loop now accelerates it in the positive z direction. Note that except for the region near the loop, the speed of the proton is nearly constant, because the electric field is weak at larger distances.

Figure 26-18*b* illustrates the motion in the case that the proton has more than enough initial kinetic energy to reach the plane of the loop. The repulsive force slows the proton's motion but does not stop it. The proton passes through the loop, with the magnitude of its velocity reaching a minimum as it passes through the loop. Once again, far from the loop the proton moves with very nearly constant velocity. ■

26-7 A DIPOLE IN AN ELECTRIC FIELD

When we place an electric dipole (Fig. 26-5*a*) in an *external* electric field, the force on the positive charge will be in one direction and the force on the negative charge will be in another direction. In accounting for the net effect of these forces, it is convenient to use the dipole moment vector $\vec{\mathbf{p}}$, which has magnitude $p = qd$ and which points from the negative charge toward the positive charge. Writing the dipole moment as a vector permits us to write the fundamental relationships involving electric dipoles in a concise form.

Figure 26-19*a* shows a dipole in a uniform electric field $\vec{\mathbf{E}}$. (This field is *not* that of the dipole itself but is produced by an external agent not shown in the figure.) The dipole moment $\vec{\mathbf{p}}$ makes an angle θ with the direction of the field. We assume the field to be uniform, so that $\vec{\mathbf{E}}$ has the same magnitude and direction at the location of $+q$ and $-q$. The forces on $+q$ and $-q$ therefore have equal magnitudes $F = qE$ but opposite directions, as shown in Fig. 26-19*a*. The net force on the dipole due to the external field is therefore zero, but there is a net torque about its center of mass that tends to rotate the dipole to bring $\vec{\mathbf{p}}$ into alignment with $\vec{\mathbf{E}}$. The torque on each charge is given by $\tau = Fr_\perp$; the net torque about the center of the dipole due to the two forces has magnitude

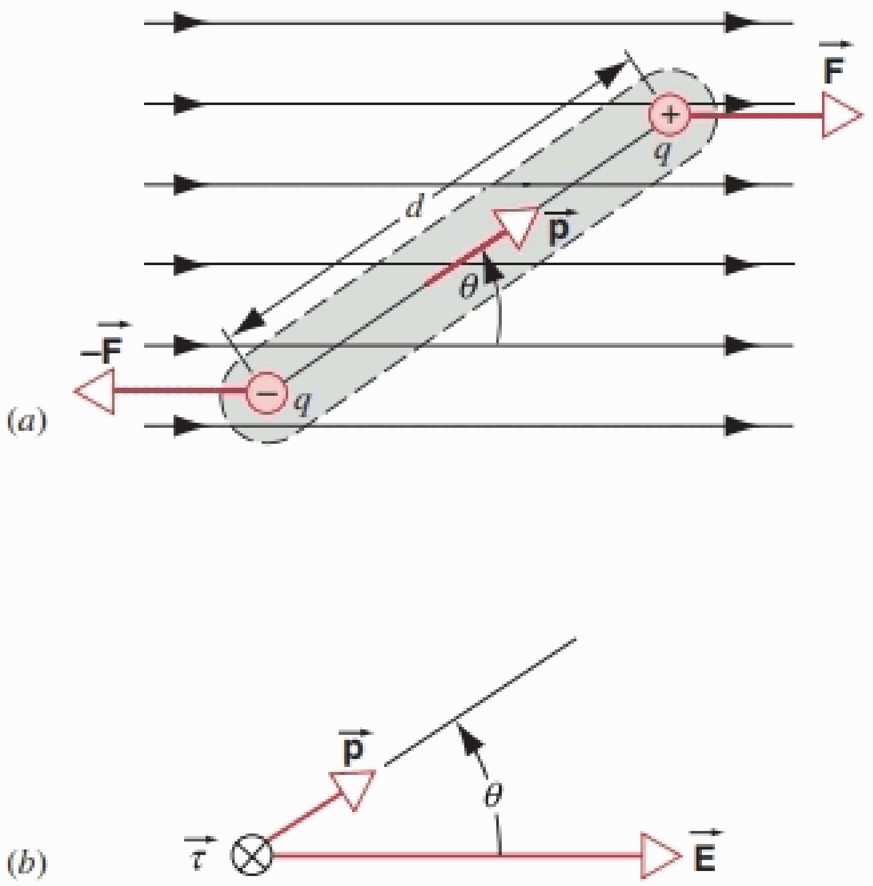


FIGURE 26-19. (*a*) An electric dipole in a uniform electric field. (*b*) The vector relationship $\vec{\boldsymbol{\tau}} = \vec{\mathbf{p}} \times \vec{\mathbf{E}}$ between the dipole moment $\vec{\mathbf{p}}$, the electric field $\vec{\mathbf{E}}$, and the resultant torque $\vec{\boldsymbol{\tau}}$ on the dipole. The torque points into the page.

$$\tau = F\frac{d}{2}\sin\theta + F\frac{d}{2}\sin\theta = Fd\sin\theta, \quad (26\text{-}25)$$

and its direction is perpendicular to the plane of the page and into the page, as indicated in Fig. 26-19*b*. We can write Eq. 26-25 as

$$\tau = (qE)d\sin\theta = (qd)E\sin\theta = pE\sin\theta. \quad (26\text{-}26)$$

Equation 26-26 can be written in vector form as

$$\vec{\boldsymbol{\tau}} = \vec{\mathbf{p}} \times \vec{\mathbf{E}}, \quad (26\text{-}27)$$

which is consistent with the directional relationships for the cross product, as shown by the three vectors in Fig. 26-19*b*.

As is generally the case in dynamics when conservative forces act (the electrostatic force is conservative, as we discuss in Chapter 28), we can represent the system equally well using either force equations or energy equations. Let us therefore consider the work done by the electric field in turning the dipole through an angle θ. Using the appropriate expression for work in rotational motion (Eq. 11-25), the work done by the external field in turning the dipole from an initial angle θ_0 to a final angle θ is

$$W = \int dW = \int_{\theta_0}^{\theta} \vec{\boldsymbol{\tau}} \cdot d\vec{\boldsymbol{\theta}} = \int_{\theta_0}^{\theta} -\tau\, d\theta, \quad (26\text{-}28)$$

where $\vec{\tau}$ is the torque exerted by the external electric field. The minus sign in Eq. 26-28 is necessary because the torque τ tends to *decrease* θ; in vector terminology, $\vec{\tau}$ and $d\vec{\theta}$ are in opposite directions, so $\vec{\tau}\cdot d\vec{\theta} = -\tau\, d\theta$. Combining Eq. 26-28 with Eq. 26-26, we obtain

$$W = \int_{\theta_0}^{\theta} -pE\sin\theta\, d\theta = -pE\int_{\theta_0}^{\theta}\sin\theta\, d\theta$$
$$= pE(\cos\theta - \cos\theta_0). \qquad (26\text{-}29)$$

Since the work done by the agent that produces the external field is equal to the negative of the change in potential energy of the system of field + dipole, we have

$$\Delta U \equiv U(\theta) - U(\theta_0) = -W = -pE(\cos\theta - \cos\theta_0). \qquad (26\text{-}30)$$

We arbitrarily define the reference angle θ_0 to be 90° and choose the potential energy $U(\theta_0)$ to be zero at that angle. At any angle θ the potential energy is then

$$U = -pE\cos\theta, \qquad (26\text{-}31)$$

which can be written in vector form as

$$U = -\vec{\mathbf{p}}\cdot\vec{\mathbf{E}}. \qquad (26\text{-}32)$$

Thus U is a minimum when $\vec{\mathbf{p}}$ and $\vec{\mathbf{E}}$ are parallel.

A water molecule has an electric dipole moment. In a microwave oven, the electric field of the microwave radiation tends to rotate the dipole moment of water molecules into alignment with the field. A free water molecule would simply oscillate back and forth about its equilibrium position, but in materials (such as food) interactions between neighboring water molecules convert the angular motion due to the torque (or, equivalently, the kinetic energy of rotation arising from the decreased potential energy of the dipole in the field) into internal energy. The direction of the electric field reverses every 2×10^{-10} s, and as the dipole moments continually try to follow the field they transfer energy that cooks the food.

We can interpret the motion of a dipole in an external field either on the basis of a torque that rotates the dipole into alignment with the field (Eq. 26-27) or a potential energy that becomes a minimum when the dipole is aligned with the field (Eq. 26-32). The choice between the two interpretations is usually based on convenience in applications to a particular problem.

SAMPLE PROBLEM 26-7. A molecule of water vapor (H_2O) has an electric dipole moment of magnitude $p = 6.2 \times 10^{-30}$ C·m. (This large dipole moment is responsible for many of the properties that make water such an important substance, such as its ability to act as an almost universal solvent.) Figure 26-20 is a representation of this molecule, showing the three nuclei and the surrounding electron distributions. The electric dipole moment $\vec{\mathbf{p}}$ is represented by a vector on the axis of symmetry. The dipole moment arises because the effective center of positive charge does not coincide with the effective center of negative charge. (A contrasting case is that of a molecule of carbon dioxide, CO_2. Here the three atoms are joined in a straight line, with a carbon in the middle and oxygens on either side. The center of positive charge and the center of negative charge coincide at the center of mass of the molecule, and the electric dipole moment of CO_2 is zero.) (*a*) How far apart are the effective centers of positive and negative charge in a molecule of H_2O? (*b*) What is the maximum torque on a molecule of H_2O in a typical laboratory electric field of magnitude 1.5×10^4 N/C? (*c*) Suppose the dipole moment of a molecule of H_2O is initially pointing in a direction opposite to the field. How much work is done by the electric field in rotating the molecule into alignment with the field?

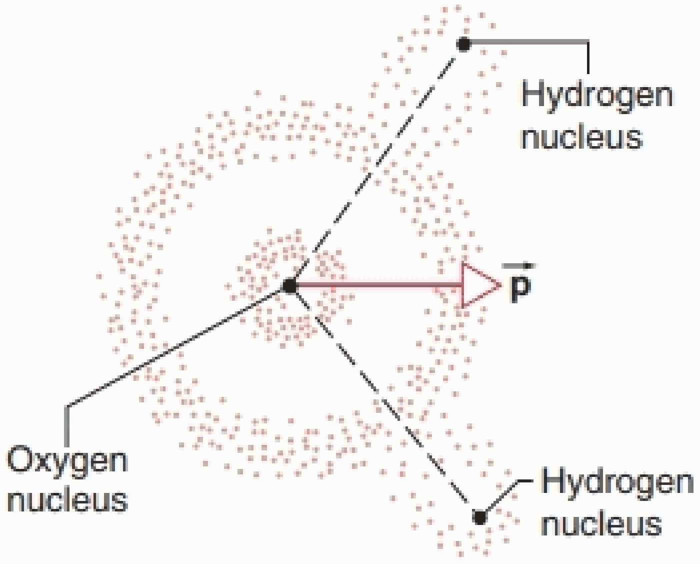


FIGURE 26-20. A molecule of H_2O, showing the three nuclei, the electron distributions, and the electric dipole moment vector $\vec{\mathbf{p}}$.

Solution (*a*) There are 10 electrons and, correspondingly, 10 positive charges in this molecule. We can write, for the magnitude of the dipole moment,

$$p = qd = (10e)(d),$$

in which d is the separation we are seeking and e is the elementary charge. Thus

$$d = \frac{p}{10e} = \frac{6.2 \times 10^{-30}\ \mathrm{C\cdot m}}{(10)(1.60 \times 10^{-19}\ \mathrm{C})}$$
$$= 3.9 \times 10^{-12}\ \mathrm{m} = 3.9\ \mathrm{pm}.$$

This is about 4% of the OH bond distance in this molecule.

(*b*) As Eq. 26-26 shows, the torque is a maximum when $\theta = 90°$. Substituting this value in that equation yields

$$\tau = pE\sin\theta = (6.2 \times 10^{-30}\ \mathrm{C\cdot m})(1.5 \times 10^4\ \mathrm{N/C})(\sin 90°)$$
$$= 9.3 \times 10^{-26}\ \mathrm{N\cdot m}.$$

(*c*) The work done in rotating the dipole from $\theta_0 = 180°$ to $\theta = 0°$ is given by Eq. 26-29,

$$\begin{aligned} W &= pE(\cos\theta - \cos\theta_0) \\ &= pE(\cos 0° - \cos 180°) \\ &= 2pE = (2)(6.2 \times 10^{-30}\ \mathrm{C\cdot m})(1.5 \times 10^4\ \mathrm{N/C}) \\ &= 1.9 \times 10^{-25}\ \mathrm{J}. \end{aligned}$$

By comparison, the average translational contribution to the internal energy ($=\frac{3}{2}kT$) of a molecule at room temperature is 6.2×10^{-21} J, which is 33,000 times larger. For the conditions of this problem, thermal agitation would overwhelm the tendency of the dipoles to align themselves with the field. That is, if we had a collection of molecules at room temperature with randomly oriented dipole moments, the application of an electric field of this magnitude would have a negligible influence on aligning the dipole moments, because of the large internal energies. If we wish to align the dipoles, we must use much stronger fields and/or much lower temperatures.

26-8 THE NUCLEAR MODEL OF THE ATOM (Optional)

Today we know that an atom consists of a tiny nucleus carrying a positive charge Ze, where Z is the atomic number of the atom. The nucleus is surrounded by a much larger volume containing Z electrons, each carrying a charge of $-e$, so that the atom as a whole is electrically neutral. We also know that the nucleus contains a very large fraction (typically greater than 99.995%) of the mass of the atom.

In the early years of the 20th century these facts were not known, and there was much speculation about the structure of the atom and especially about the distribution of its positive charge. According to one theory that was popular at that time, the positive charge is distributed more or less uniformly throughout the entire spherical volume of the atom. This model of the structure of the atom is called the *Thomson model* after J. J. Thomson, who proposed it. (Thomson was the first to measure the charge-to-mass ratio of the electron and is therefore often credited as the discoverer of the electron.) It is also called the "plum pudding" model, because the electrons are imbedded throughout the diffuse sphere of positive charge like raisins in a plum pudding.

One way of testing this model is to determine the electric field of the atom by probing it with a beam of positively charged projectiles that pass nearby. The particles in the beam are deflected or *scattered* by the electric field of the atom. In the following discussion, we consider only the effect on the projectile of the sphere of positive charge. We assume that the projectile is both much *less* massive than the atom and much *more* massive than an electron. In this way the electrons have a negligible effect on the scattering of the projectile, and the atom can be assumed to remain at rest while the projectile is deflected.

We can estimate the deflection for a Thomson-model atom, in which the positive charge is uniformly distributed throughout the volume of the atom. The electric field due to a uniform sphere of positive charge was given by Eq. 26-6 for points outside the sphere of charge and by Eq. 26-24 for points inside. Let us calculate the electric field at the surface, which, as Fig. 26-8 shows, is the *largest* possible field that this distribution can produce. We consider a heavy atom such as gold, which has a positive charge Q of $79e$ and a radius R of about 1.0×10^{-10} m. Neglecting the electrons, the electric field at $r = R$ due to the positive charges is

$$E_{\text{max}} = \frac{1}{4\pi\epsilon_0}\frac{Q}{R^2} = 1.1 \times 10^{13}\ \text{N/C}.$$

For the projectiles in our experiment, let us use a beam of alpha particles, which have a positive charge q of $2e$ and a mass m of 6.6×10^{-27} kg. Alpha particles are nuclei of helium atoms, which are emitted in certain radioactive decay processes. A typical kinetic energy for such a particle might be about $K = 6$ MeV or 9.6×10^{-13} J. At this energy you can easily verify that the particle has a speed of about 1.7×10^7 m/s.

Let the particle pass near the surface of the atom, where it experiences the largest electric field that this atom could exert. The corresponding force on the particle is

$$F = qE_{\text{max}} = 3.5 \times 10^{-6}\ \text{N}.$$

Figure 26-21 shows a schematic diagram of a scattering experiment. The actual calculation of the deflection is relatively complicated, but we can make some approximations that simplify the calculation and permit a rough estimate of the maximum deflection. Let us assume that the above force is constant and acts only during the time Δt it takes the projectile to travel a distance equal to a diameter of the atom, as indicated in Fig. 26-21. This time interval is

$$\Delta t = \frac{2R}{v} = 1.2 \times 10^{-17}\ \text{s}.$$

The force gives the particle a transverse acceleration a, which produces a transverse velocity Δv given by

$$\Delta v = a\,\Delta t = \frac{F}{m}\,\Delta t = 6.4 \times 10^3\ \text{m/s}.$$

The particle will be deflected by a small angle θ that can be estimated to be about

$$\theta = \tan^{-1}\frac{\Delta v}{v} = 0.02°.$$

This type of scattering experiment was first done by Ernest Rutherford and his collaborators at the University of

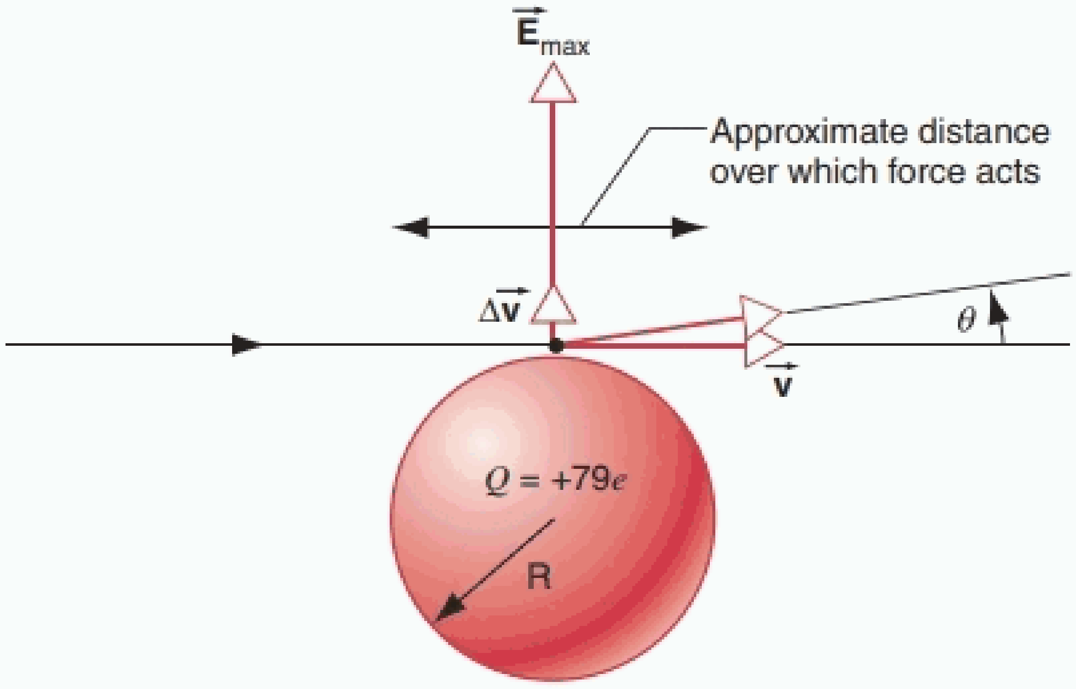


FIGURE 26-21. The scattering of a positively charged projectile passing near the surface of an atom, represented by a uniform sphere of positive charge. The electric field on the projectile causes a transverse deflection by an angle θ.

Manchester (England) in 1911. They passed a beam of alpha particles through a thin gold foil and determined the relative probability for alpha particles to be scattered through various angles θ relative to their original direction. Of course they could not control how the alpha particles passed through any particular atom; in fact, rather than grazing the edge, most alpha particles would pass through the volume of the atom pictured in Fig. 26-21 and (according to the Thomson model) be deflected by less than the maximum angle we have calculated.

The results of the experiment showed that, although nearly all of the alpha particles were deflected by angles no greater than a few hundredths of a degree, a small number (perhaps 1 in 10^4) were deflected by angles greater than 90°. This result is in complete disagreement with the Thomson model and led Rutherford to comment: "It was quite the most incredible event that ever happened to me in my life. It was almost as incredible as if you had fired a 15-inch shell at a piece of tissue paper and it came back and hit you."

Based on this kind of scattering experiment, Rutherford concluded that the positive charge of an atom was *not* diffused throughout a sphere of the same size as the atom, but instead was concentrated in a tiny region (the *nucleus*) near the center of the atom. In the case of a gold atom, the nucleus has a radius of about 7×10^{-15} m (7 fm), roughly 10^{-4} times smaller than the radius of the atom. That is, the nucleus occupies a volume only 10^{-12} that of the atom!

Let us calculate the maximum electric field and the corresponding force on an alpha particle that passes close to the surface of the nucleus. If we regard the nucleus as a uniform spherical ball of charge $Q = 79e$ and radius $R = 7$ fm, the maximum electric field is

$$E_{\max} = \frac{1}{4\pi\epsilon_0}\frac{Q}{R^2} = 2.3 \times 10^{21}\ \text{N/C}.$$

This is more than eight orders of magnitude larger than the electric field that would act on a particle at the surface of a plum-pudding model atom. The corresponding force is

$$F = qE_{\max} = 740\ \text{N}.$$

This is a huge force! Let us make the same simplification we did in our previous calculation and assume that this force is constant and acts on the particle only during the time Δt it takes the particle to travel a distance equal to one nuclear diameter:

$$\Delta t = \frac{2R}{v} = 8.2 \times 10^{-22}\ \text{s}.$$

The corresponding change in the velocity of the particle can be estimated to be

$$\Delta v = a\,\Delta t = \frac{F}{m}\Delta t = 9 \times 10^7\ \text{m/s}.$$

This is comparable in magnitude to the velocity itself. We conclude that a nuclear atom can produce an electric field that is sufficiently large to reverse the motion of the projectile.

Based on the nuclear model of the atom, Rutherford was able to derive an exact formula for the number of particles scattered at any particular angle, and the experiments showed perfect agreement with this formula. He was also able to use the formula to determine the atomic number Z of the target atoms. Moreover, by using particles of a higher energy that actually penetrate the nucleus (see Sample Problem 25-7), this method can also be used to determine the nuclear radius.

This classic and painstaking series of experiments and their brilliant interpretation laid the foundation for modern atomic and nuclear physics, and Rutherford is generally credited as the founder of these fields. ■

MULTIPLE CHOICE

26-1 What Is a Field?

26-2 The Electric Field

1. The electric field is defined in Eq. 26-3 in terms of q_0, a small *positive* charge. If instead the definition were in terms of a small *negative* charge of the same magnitude, then compared to the *original* field, the newly defined electric field

- (A) would point in the same direction, and have the same magnitude.
- (B) would point in the opposite direction, but have the same magnitude.
- (C) would point in the same direction, but have a different magnitude.
- (D) would point in the opposite direction, and have a different magnitude.

26-3 The Electric Field of a Point Charge

2. A point charge $+q$ is located at the origin, and a point charge $+2q$ is located at $x = a$, where a is positive.

(*a*) Which of the following statements is true?

- (A) Close to the charges, the electric field can be zero off the x axis.
- (B) Close to the charges, the magnitude of the electric field can be a maximum off the x axis.
- (C) The electric field can be zero somewhere between the charges.
- (D) The electric field can be zero on the x axis at finite points not between the charges.

(*b*) In which of the following regions might there exist a point where the electric field has zero magnitude?

(A) $-\infty < x < 0$ (B) $0 < x < a$
(C) $a < x < \infty$
(D) E does not vanish in the region $-\infty < x < \infty$.

3. A point charge $+q$ is located at the origin, and a point charge $-2q$ is located at $x = a$, where a is positive.

(*a*) Which of the following statements is true?
(A) Close to the charges, the electric field can be zero off the x axis.
(B) Close to the charges, the magnitude of the electric field can be a maximum off the x axis.
(C) The electric field can be zero between the charges.
(D) The electric field can be zero on the x axis at finite points not between the charges.

(*b*) In which of the following regions might there exist a point where the electric field has zero magnitude?
(A) $-\infty < x < 0$ (B) $0 < x < a$
(C) $a < x < \infty$
(D) E does not vanish in the region $-\infty < x < \infty$.

26-4 Electric Field of Continuous Charge Distributions

4. Consider the magnitude of the electric field $E(z)$ on the axis of a uniform ring of charge.

(*a*) $E(z)$ will have its largest value where
(A) $z = 0$. (B) $0 < |z| < \infty$.
(C) $|z| = \infty$. (D) (A) and (C) are correct.

(*b*) $E(z)$ can be zero where
(A) $z = 0$. (B) $0 < |z| < \infty$.
(C) $|z| = \infty$. (D) (A) and (C) are correct.

5. Consider the magnitude of the electric field $E(z)$ on the axis of a uniform disk of charge.

(*a*) $E(z)$ will have its largest value where
(A) $z = 0$. (B) $0 < |z| < \infty$.
(C) $|z| = \infty$. (D) (A) and (C) are correct.

(*b*) $E(z)$ can be zero where
(A) $z = 0$. (B) $0 < |z| < \infty$.
(C) $|z| = \infty$. (D) (A) and (C) are correct.

26-5 Electric Field Lines

6. Figure 26-22 shows the electric field lines around an electric dipole. Which of the arrows best represents the electric field at point P?

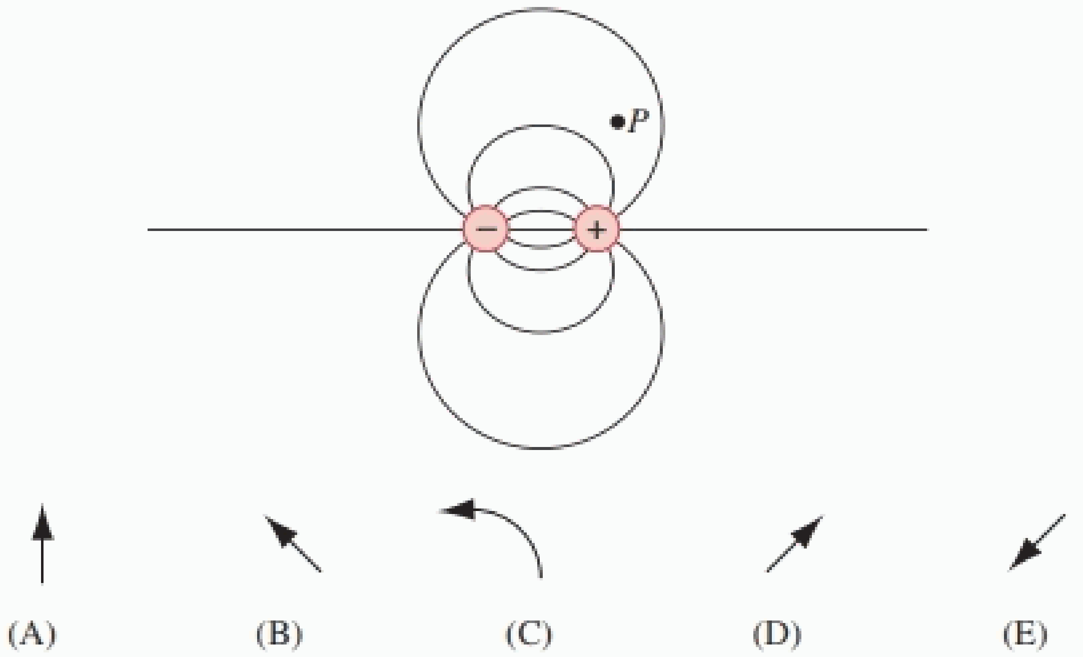


FIGURE 26-22. Multiple-choice question 6.

7. Figure 26-23 shows the electric field lines around three point charges, A, B, and C. (*a*) Which charges are positive? (*b*) Which charge has the largest magnitude? (*c*) In which region or regions of the picture could the electric field be zero?
(A) near A (B) near B
(C) near C (D) nowhere

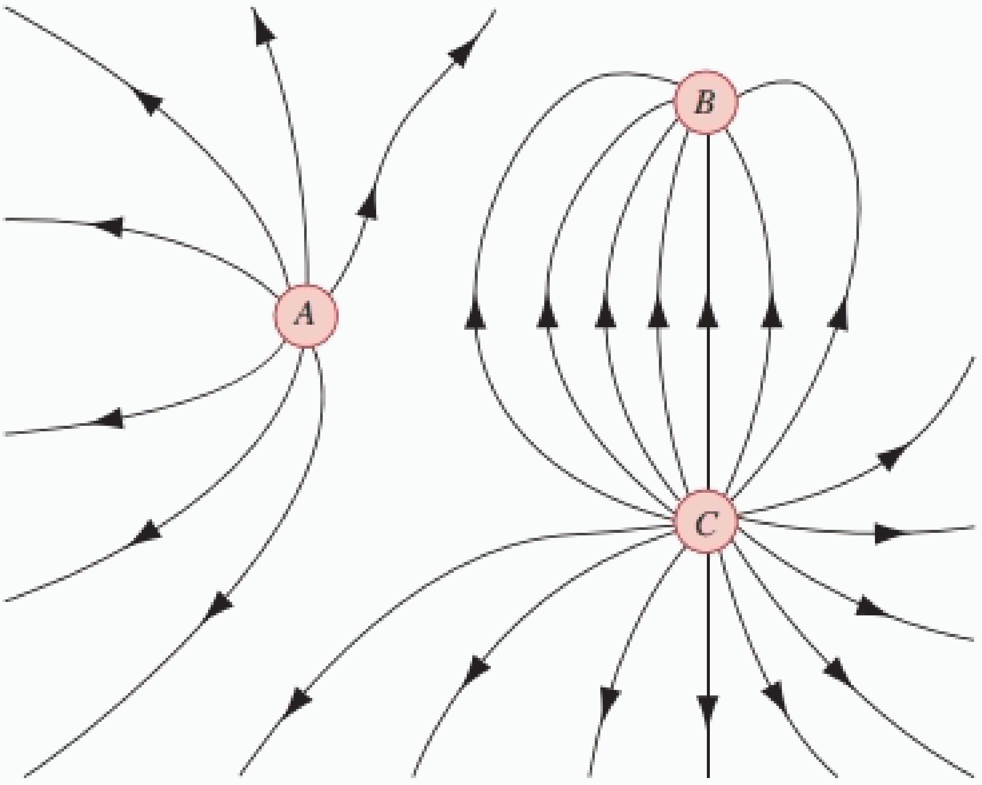


FIGURE 26-23. Multiple-choice question 7.

26-6 A Point Charge in an Electric Field

8. Three small spheres x, y, and z carry charges of equal magnitudes and with signs shown in Fig. 26-24. They are placed at the vertices of an isosceles triangle with the distance between x and y equal to the distance between x and z. Spheres y and z are held in place but sphere x is free to move on a frictionless surface.

(*a*) What is the direction of the electric force on sphere x at the point shown in the figure?

(*b*) Which path will sphere x take when released?

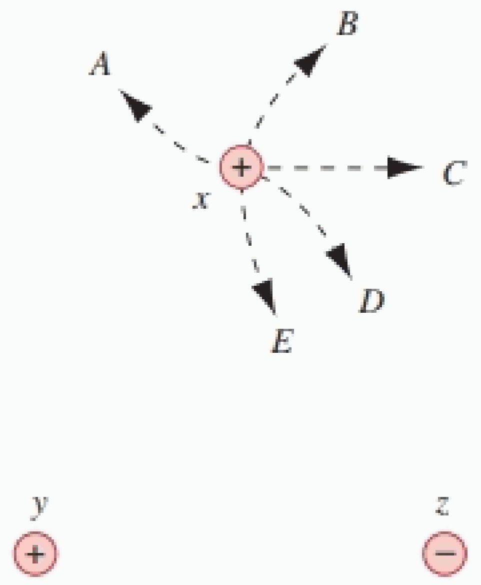


FIGURE 26-24. Multiple-choice question 8.

9. An electron is located in the *uniform* electric field established between parallel charged positive and negative plates. Where would the electron experience the greatest electrostatic force?
(A) When the electron is closer to the positive plate
(B) When the electron is closer to the negative plate
(C) When the electron is midway between the plates
(D) The electron experiences the same force regardless of the location between the plates.

10. The following measurements were made of the charge (in units of 10^{-19} C) on a series of charged droplets. What is the largest possible fundamental unit of charge that can be deduced from these data?

48	19.2	28.8
9.6	38.4	24

(A) 1.6×10^{-19} C (B) 4.8×10^{-19} C
(C) 9.6×10^{-19} C (D) 48×10^{-19} C

26-7 A Dipole in an Electric Field

11. The electric field in a certain region of space obeys $E_y \neq 0$, $E_x = E_z = 0$ and $\partial\vec{\mathbf{E}}/\partial x \neq 0$, $\partial\vec{\mathbf{E}}/\partial y = \partial\vec{\mathbf{E}}/\partial z = 0$.

(*a*) The net force on an electric dipole oriented parallel to the x axis in this field is

(A) directed along the x axis.
(B) directed along the y axis.
(C) directed along the z axis.
(D) None of the above

(*b*) The net torque on an electric dipole parallel to the x axis in this field is

(A) directed along the x axis.
(B) directed along the y axis.
(C) directed along the z axis.
(D) None of the above

26-8 The Nuclear Model of the Atom

QUESTIONS

1. Name as many scalar fields and vector fields as you can.

2. (*a*) In the gravitational attraction between the Earth and a stone, can we say that the Earth lies in the gravitational field of the stone? (*b*) How is the gravitational field due to the stone related to that due to the Earth?

3. A positively charged ball hangs from a long silk thread. We wish to measure E at a point in the same horizontal plane as that of the hanging charge. To do so, we put a positive test charge q_0 at the point and measure F/q_0. Will F/q_0 be less than, equal to, or greater than E at the point in question?

4. In exploring electric fields with a test charge, we have often assumed, for convenience, that the test charge was positive. Does this really make any difference in determining the field? Illustrate in a simple case of your own devising.

5. Electric field lines never cross. Why?

6. In Fig. 26-13, why do the field lines around the edge of the figure appear, when extended backward, to radiate uniformly from the center of the figure?

7. A point charge is moving in an electric field at right angles to the field lines. Does any force act on it?

8. In Fig. 26-14, why should grass seeds line up with electric field lines? Grass seeds normally carry no electric charge. (See "Demonstration of the Electric Fields of Current-Carrying Conductors," by O. Jefimenko, *American Journal of Physics,* January 1962, p. 19.)

9. What is the origin of "static cling," a phenomenon that sometimes affects clothes as they are removed from a dryer?

10. Two point charges of unknown magnitude and sign are a distance d apart. The electric field is zero at one point between them, on the line joining them. What can you conclude about the charges?

11. Two point charges of unknown magnitude and sign are placed a distance d apart. (*a*) If it is possible to have $E = 0$ at any point not between the charges but on the line joining them, what are the necessary conditions and where is the point located? (*b*) Is it possible, for any arrangement of two points charges, to find two points (neither at infinity) at which $E = 0$? If so, under what conditions?

12. Two point charges of unknown sign and magnitude are fixed a distance d apart. Can we have $E = 0$ for off-axis points (excluding infinity)? Explain.

13. In Sample Problem 26-3, a charge placed at point P in Fig. 26-4 is in equilibrium because no force acts on it. Is the equilibrium stable (*a*) for displacements along the line joining the charges and (*b*) for displacements at right angles to this line?

14. In Fig. 26-12, the force on the lower charge points up and is finite. The crowding of the field lines, however, suggests that E is infinitely great at the site of this (point) charge. A charge immersed in an infinitely great field should have an infinitely great force acting on it. What is the solution to this dilemma?

15. A point charge q of mass m is released from rest in a nonuniform field. (*a*) Will it necessarily follow the electric field line that passes through the release point? (*b*) Under what circumstances, if any, will a charged particle follow the electric field lines?

16. A positive and a negative charge of the same magnitude lie on a long straight line. What is the direction of $\vec{\mathbf{E}}$ for points on this line that lie (*a*) between the charges, (*b*) outside the charges in the direction of the positive charge, (*c*) outside the charges in the direction of the negative charge, and (*d*) off the line but in the median plane of the charges?

17. In the median plane of an electric dipole, is the electric field parallel or antiparallel to the electric dipole moment $\vec{\mathbf{p}}$?

18. In what way does Eq. 26-12 fail to represent the field lines of Fig. 26-12 if we relax the requirement that $x \gg d$?

19. (*a*) Two identical electric dipoles are placed in a straight line, as shown in Fig. 26-25*a*. What is the direction of the electric force on each dipole due to the presence of the other? (*b*) Suppose that the dipoles are rearranged as in Fig. 26-25*b*. What now is the direction of the force?

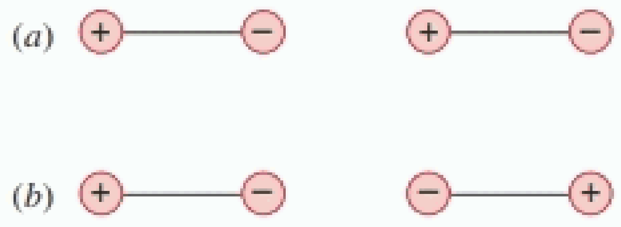


FIGURE 26-25. Question 19.

20. Compare the way E varies with r for (*a*) a point charge, (*b*) a dipole, and (*c*) a quadrupole.

21. What mathematical difficulties would you encounter if you were to calculate the electric field of a charged ring (or disk) at points not on the axis?

22. Equation 26-20 shows that E_z has the same value for all points in front of an infinite uniformly charged sheet. Is this reasonable? One might think that the field should be stronger near the sheet because the charges are so much closer.

23. Describe, in your own words, the purpose of the Millikan oil-drop experiment.

24. How does the sign of the charge on the oil drop affect the operation of the Millikan experiment?

25. Why did Millikan not try to balance electrons in his apparatus instead of oil drops?

26. You turn an electric dipole end for end in a uniform electric field. How does the work you do depend on the initial orientation of the dipole with respect to the field?

27. For what orientations of an electric dipole in a uniform electric field is the potential energy of the dipole (*a*) the greatest and (*b*) the least?

28. An electric dipole is placed in a nonuniform electric field. Is there a net force on the dipole?

29. An electric dipole is placed at rest in a uniform external electric field, as in Fig. 26-19*a*, and released. Discuss its motion.

30. An electric dipole has its dipole moment $\vec{\mathbf{p}}$ aligned with a uniform external electric field $\vec{\mathbf{E}}$. (*a*) Is the equilibrium stable or unstable? (*b*) Discuss the nature of the equilibrium if $\vec{\mathbf{p}}$ and $\vec{\mathbf{E}}$ point in opposite directions.

31. An atom is normally electrically neutral. Why then should an alpha particle be deflected by the atom under any circumstances?

EXERCISES

26-1 What Is a Field?

26-2 The Electric Field

1. An electron is accelerated eastward at $1.84 \times 10^9\ \mathrm{m/s^2}$ by an electric field. Determine the magnitude and direction of the electric field.

2. Humid air breaks down (its molecules become ionized) in an electric field of 3.0×10^6 N/C. What is the magnitude of the electric force on (*a*) an electron and (*b*) an ion (with a single electron missing) in this field?

3. An alpha particle, the nucleus of a helium atom, has a mass of 6.64×10^{-27} kg and a charge of $+2e$. What are the magnitude and direction of the electric field that will balance its weight?

4. In a uniform electric field near the surface of the Earth, a particle having a charge of -2.0×10^{-9} C is acted on by a downward electric force of 3.0×10^{-6} N. (*a*) Find the magnitude of the electric field. (*b*) What are the magnitude and direction of the electric force exerted on a proton placed in this field? (*c*) What is the gravitational force on the proton? (*d*) What is the ratio of the electric force to the gravitational force in this case?

26-3 The Electric Field of a Point Charge

5. What is the magnitude of a point charge chosen so that the electric field 75.0 cm away has the magnitude 2.30 N/C?

6. Calculate the dipole moment of an electron and a proton that are 4.30 nm apart.

7. Calculate the magnitude of the electric field, due to an electric dipole of dipole moment 3.56×10^{-29} C·m, at a point 25.4 nm away along the bisector axis.

8. Find the electric field at the center of the square of Fig. 26-26. Assume that $q = 11.8$ nC and $a = 5.20$ cm.

9. A clock face has negative point charges $-q$, $-2q$, $-3q$, . . . , $-12q$ fixed at the positions of the corresponding numerals. The clock hands do not perturb the field. At what time does the hour hand point in the same direction as the electric field at the center of the dial? (Hint: Consider diametrically opposite charges.)

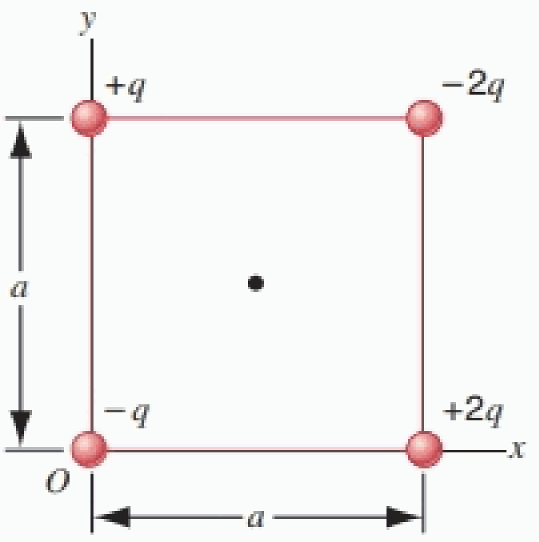


FIGURE 26-26. Exercise 8.

10. In Fig. 26-5, assume that both charges are positive. Show that the magnitude of E at point P in that figure, assuming $x \gg d$, is given by

$$E = \frac{1}{4\pi\epsilon_0}\frac{2q}{x^2}.$$

11. One type of electric quadrupole is formed by four charges located at the vertices of a square of side $2a$. Point P lies a distance x from the center of the quadrupole on a line parallel to two sides of the square as shown in Fig. 26-27. For $x \gg a$, show that the electric field at P is approximately given by

$$E = \frac{3(2qa^2)}{2\pi\epsilon_0 x^4}.$$

(Hint: Treat the quadrupole as two dipoles.)

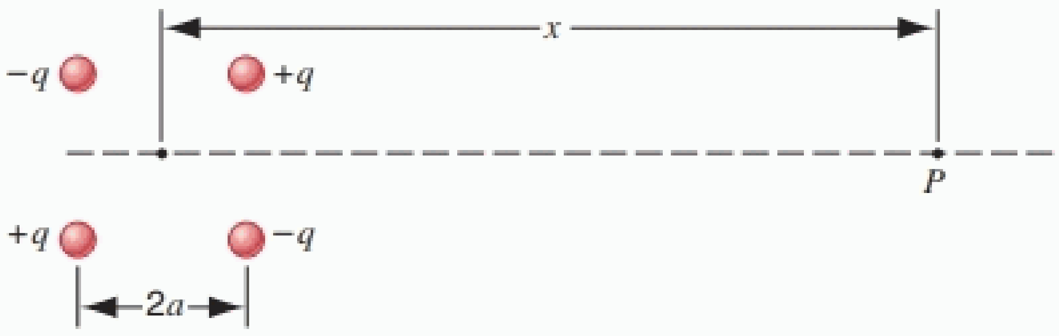


FIGURE 26-27. Exercise 11.

26-4 Electric Field of Continuous Charge Distributions

12. Show that Eq. 26-19, for the electric field of a charged disk at points on its axis, reduces to the field of a point charge for $z \gg R$.

13. At what distance along the axis of a charged disk of radius R is the electric field strength equal to one-half the value of the field at the surface of the disk at the center?

14. At what distance along the axis of a charged ring of radius R is the axial electric field strength a maximum?

15. (*a*) What total charge q must a disk of radius 2.50 cm carry so that the electric field on the surface of the disk at its center equals the value at which air breaks down electrically, producing sparks? See Table 26-1. (*b*) Suppose that each atom at the surface has an effective cross-sectional area of 0.015 nm^2. How many atoms are at the disk's surface? (*c*) The charge in (*a*) results from some of the surface atoms carrying one excess electron. What fraction of the surface atoms must be so charged?

16. A thin glass rod is bent into a semicircle of radius r. A charge $+q$ is uniformly distributed along the upper half and a charge $-q$ is uniformly distributed along the lower half, as shown in Fig. 26-28. Find the electric field $\vec{\mathbf{E}}$ at P, the center of the semicircle.

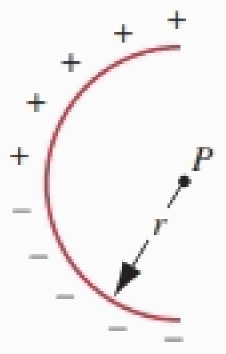


FIGURE 26-28. Exercise 16.

17. Measured values of the electric field E a distance z along the axis of a charged plastic disk are given here:

z (cm)	E (10^7 N/C)
0	2.043
1	1.732
2	1.442
3	1.187
4	0.972
5	0.797

Calculate (*a*) the radius of the disk and (*b*) the charge on it.

18. An insulating rod of length L has charge $-q$ uniformly distributed along its length, as shown in Fig. 26-29. (*a*) What is the linear charge density of the rod? (*b*) Find the electric field at point P a distance a from the end of the rod. (*c*) If P were very far from the rod compared to L, the rod would look like a point charge. Show that your answer to (*b*) reduces to the electric field of a point charge for $a \gg L$.

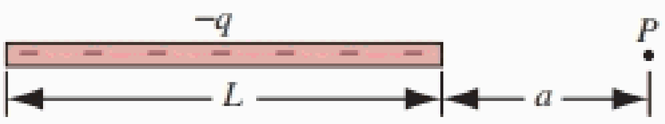


FIGURE 26-29. Exercise 18.

19. Sketch qualitatively the field lines associated with three long parallel lines of charge in a perpendicular plane. Assume that the intersections of the lines of charge with such a plane form an equilateral triangle (Fig. 26-30) and that each line of charge has the same linear charge density λ.

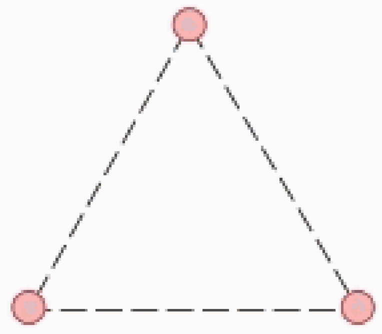


FIGURE 26-30. Exercise 19.

26-5 Electric Field Lines

20. Figure 26-31 shows field lines of an electric field; the line spacing perpendicular to the page is the same everywhere. (*a*) If the magnitude of the field at A is 40 N/C, what force does an electron at that point experience? (*fb*) What is the magnitude of the field at B?

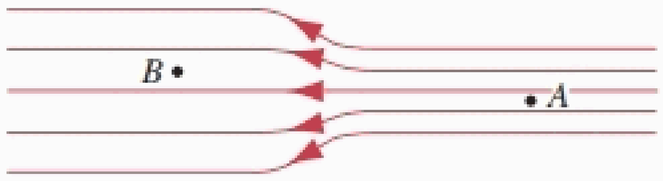


FIGURE 26-31. Exercise 20.

21. Sketch qualitatively the field lines associated with a thin, circular, uniformly charged disk of radius R. (Hint: Consider as limiting cases points very close to the disk, where the electric field is perpendicular to the surface, and points very far from it, where the electric field is like that of a point charge.)

22. Sketch qualitatively the field lines associated with two separated point charges $+q$ and $-2q$.

23. Three charges are arranged in an equilateral triangle as in Fig. 26-32. Consider the field lines due to $+Q$ and $-Q$, and from them identify the direction of the force that acts on $+q$ because of the presence of the other two charges. (Hint: See Fig. 26-12.)

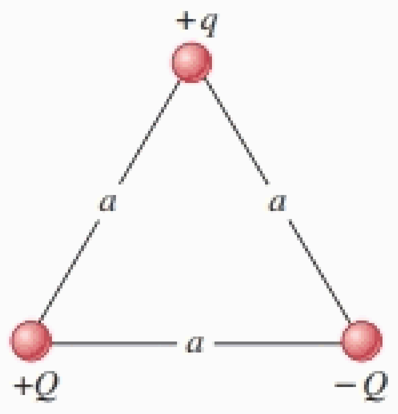


FIGURE 26-32. Exercise 23.

24. (*a*) In Fig. 26-33, locate the point (or points) at which the electric field is zero. (*b*) Sketch qualitatively the field lines.

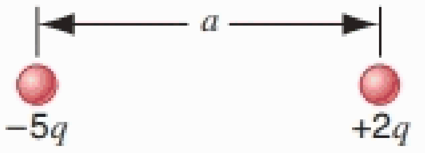


FIGURE 26-33. Exercise 24.

25. Two point charges are fixed at a distance d apart on the x axis (Fig. 26-34). Plot $E_x(x)$, assuming $x = 0$ at the left-hand charge. Consider both positive and negative values of x. Plot E_x as positive if $\vec{\mathbf{E}}$ points to the right and negative if $\vec{\mathbf{E}}$ points to the left. Assume $q_1 = +1.0 \times 10^{-6}$ C, $q_2 = +3.0 \times 10^{-6}$, and $d = 10$ cm.

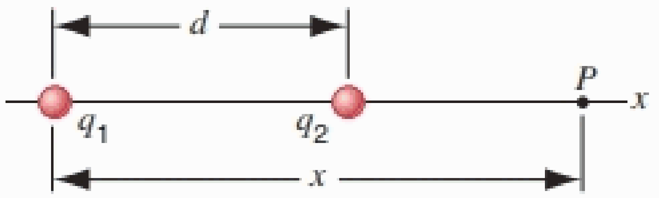


FIGURE 26-34. Exercise 25.

26. Charges $+q$ and $-2q$ are fixed a distance d apart as in Fig. 26-35. (*a*) Find $\vec{\mathbf{E}}$ at points A, B, and C. (*b*) Sketch roughly the electric field lines.

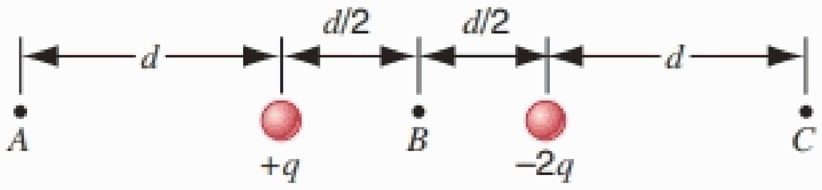


FIGURE 26-35. Exercise 26.

26-6 A Point Charge in an Electric Field

27. An electron moving with a speed of 4.86×10^6 m/s is shot parallel to a uniform electric field of strength 1030 N/C arranged so as to retard its motion. (*a*) How far will the electron travel in the field before coming (momentarily) to rest and (*b*) how much time will elapse? (*c*) If the electric field ends abruptly after 7.88 mm, what fraction of its initial kinetic energy will the electron lose in traversing it?

28. One weapon being considered for anti-missile defense uses particle beams. For example, a proton beam striking an enemy missile could render it harmless. Such beams can be produced in "guns" using electric fields to accelerate the charged particles. (*a*) What acceleration would a proton experience if the electric field is 2.16×10^4 N/C? (*b*) What speed would the proton attain if the field acts over a distance of 1.22 cm?

29. Two equal and opposite charges of magnitude 1.88×10^{-7} C are held 15.2 cm apart. (*a*) What are the magnitude and direction of $\vec{\mathbf{E}}$ at a point midway between the charges? (*b*) What force (magnitude and direction) would act on an electron placed there?

30. A uniform electric field exists in a region between two oppositely charged plates. An electron is released from rest at the surface of the negatively charged plate and strikes the surface of the opposite plate, 1.95 cm away, 14.7 ns later. (*a*) What is the speed of the electron as it strikes the second plate? (*b*) What is the magnitude of the electric field?

31. In Millikan's experiment, a drop of radius 1.64 μm and density 0.851 g/cm^3 is balanced when an electric field of 1.92×10^5 N/C is applied. Find the charge on the drop, in terms of e.

32. Two point charges of magnitudes $q_1 = 2.16\ \mu$C and $q_2 = 85.3$ nC are 11.7 cm apart. (*a*) Find the magnitude of the electric field that each produces at the site of the other. (*b*) Find the magnitude of the force on each charge.

33. In a particular early run (1911), Millikan observed that the following measured charges, among others, appeared at different times on a single drop:

6.563×10^{-19} C	13.13×10^{-19} C	19.71×10^{-19} C
8.204×10^{-19} C	16.48×10^{-19} C	22.89×10^{-19} C
11.50×10^{-19} C	18.08×10^{-19} C	26.13×10^{-19} C

What value for the quantum of charge e can be deduced from these data?

34. A uniform vertical field $\vec{\mathbf{E}}$ is established in the space between two large parallel plates. A small conducting sphere of mass m is suspended in the field from a string of length L. Find the period of this pendulum when the sphere is given a charge $+q$ if the lower plate (*a*) is charged positively and (*b*) is charged negatively.

35. In Sample Problem 26-6, find the total deflection of the ink drop on striking the paper 6.8 mm from the end of the deflection plates; see Fig. 26-16.

26-7 A Dipole in an Electric Field

36. An electric dipole, consisting of charges of magnitude 1.48 nC separated by 6.23 μm, is in an electric field of strength 1100 N/C. (*a*) What is the magnitude of the electric dipole moment? (*b*) What is the difference in potential energy corresponding to dipole orientations parallel and antiparallel to the field?

37. An electric dipole consists of charges $+2e$ and $-2e$ separated by 0.78 nm. It is in an electric field of strength 3.4×10^6 N/C. Calculate the magnitude of the torque on the dipole when the dipole moment is (*a*) parallel, (*b*) at a right angle, and (*c*) opposite to the electric field.

38. A charge $q = 3.16\ \mu$C is 28.5 cm from a small dipole along its perpendicular bisector. The force on the charge equals 5.22×10^{-16} N. Show on a diagram (*a*) the direction of the force on the charge and (*b*) the direction of the force on the dipole. Determine (*c*) the magnitude of the force on the dipole and (*d*) the dipole moment of the dipole.

26-8 The Nuclear Model of the Atom

39. In a 1911 paper, Ernest Rutherford said: In order to form some idea of the forces required to deflect an alpha particle through a large angle, consider an atom containing a point positive charge Ze at its center and surrounded by a distribution of negative electricity, $-Ze$ uniformly distributed within a sphere of radius R. The electric field E . . . at a distance r from the center for a point inside the atom [is]

$$E = \frac{Ze}{4\pi\epsilon_0}\left(\frac{1}{r^2} - \frac{r}{R^3}\right).$$

Verify this equation.

40. Figure 26-36 shows a Thomson atom model of helium ($Z = 2$). Two electrons, at rest, are embedded inside a uniform sphere of positive charge $2e$. Find the distance d between the electrons so that the configuration is in static equilibrium.

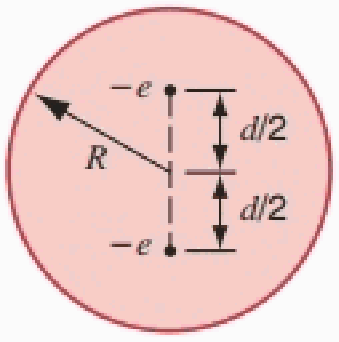


FIGURE 26-36. Exercise 40.

PROBLEMS

1. In Fig. 26-5, consider a point that is a distance z from the center of a dipole along its axis. (*a*) Show that, at large values of z, the magnitude of the electric field is given by

$$E = \frac{1}{2\pi\epsilon_0}\frac{p}{z^3}.$$

(Compare with the field at a point on the perpendicular bisector.) (*b*) What is the direction of $\vec{\mathbf{E}}$?

2. Show that the components of $\vec{\mathbf{E}}$ due to a dipole are given, at distant points, by

$$E_x = \frac{1}{4\pi\epsilon_0}\frac{3pxz}{(x^2+z^2)^{5/2}}, \qquad E_z = \frac{1}{4\pi\epsilon_0}\frac{p(2z^2-x^2)}{(x^2+z^2)^{5/2}},$$

where x and z are coordinates of point P in Fig. 26-37. Show that this general result includes the special results of Eq. 26-12 and Problem 1.

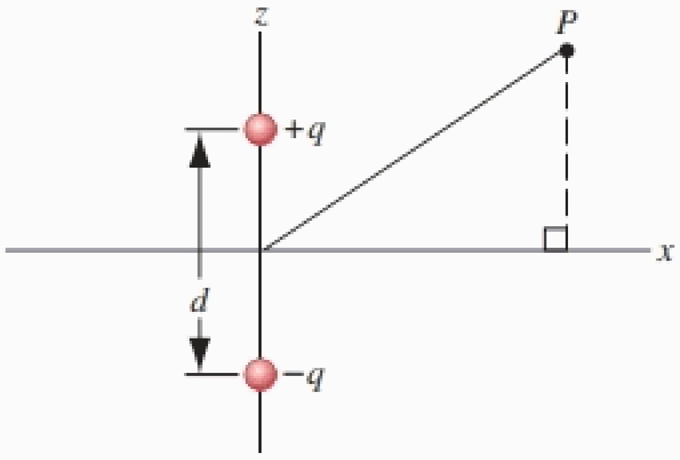


FIGURE 26-37. Problem 2.

3. Consider the ring of charge of Section 26-4. Suppose that the charge q is not distributed uniformly over the ring but that charge q_1 is distributed uniformly over half the circumference and charge q_2 is distributed uniformly over the other half. Let $q_1 + q_2 = q$. (*a*) Find the component of the electric field at any point on the axis directed along the axis and compare with the uniform case. (*b*) Find the component of the electric field at any point on the axis perpendicular to the axis and compare with the uniform case.

4. Figure 26-38 shows one type of electric quadrupole. It consists of two dipoles whose effects at external points do not quite cancel. Show that the value of E on the axis of the quadrupole for points a distance z from its center (assume $z \gg d$) is given by

$$E = \frac{3Q}{4\pi\epsilon_0 z^4},$$

where Q ($=2qd^2$) is called the *quadrupole moment* of the charge distribution.

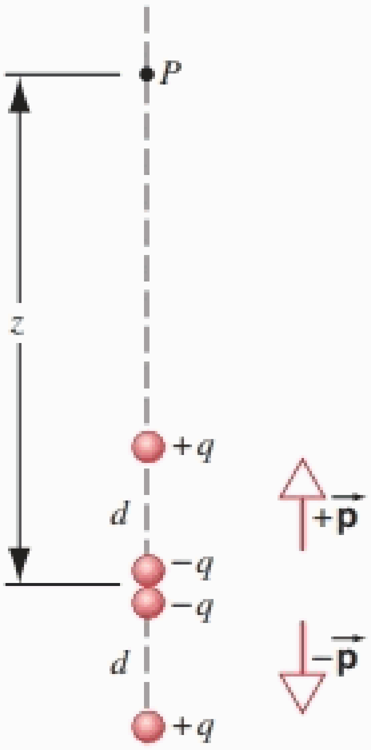


FIGURE 26-38. Problem 4.

5. Construct a distribution of point charges along the x axis so that, far from the charges, the electric field along the x axis falls off as $1/r^6$.

6. A "semi-infinite" insulating rod (Fig. 26-39) carries a constant charge per unit length of λ. Show that the electric field at the point P makes an angle of 45° with the rod and that this result is independent of the distance R.

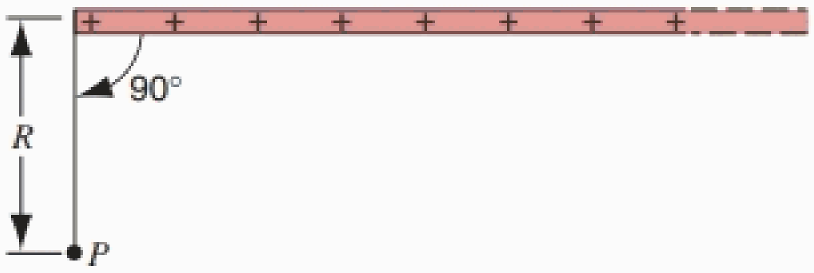


FIGURE 26-39. Problem 6.

7. A thin nonconducting rod of finite length L carries a uniform linear charge density $+\lambda$ on the top half and a uniform charge density $-\lambda$ on the bottom half; compare to Fig. 26-6. (*a*) Use a symmetry argument to determine the direction of the electric field at P due to the rod. (*b*) Find $\vec{\mathbf{E}}$ at P. (*c*) Take the limit of this expression for large y. How does it depend on y? What does this remind you of?

8. A nonconducting hemispherical cup of inner radius R has a total charge q spread uniformly over its inner surface. Find the electric field at the center of curvature. (Hint: Consider the cup as a stack of rings.)

9. Assume that the exponent in Coulomb's law is not 2 but n. Show that for $n \neq 2$ it is impossible to construct lines that will have the properties listed for electric field lines in Section 26-5. For simplicity, treat an isolated point charge.

10. Two large parallel copper plates are 5.00 cm apart and have a uniform electric field between them as depicted in Fig. 26-40. An electron is released from the negative plate at the same time that a proton is released from the positive plate. Neglect the force of the particles on each other and find their distance from the positive plate when they pass each other. Does it surprise you that you need not know the electric field to solve this problem?

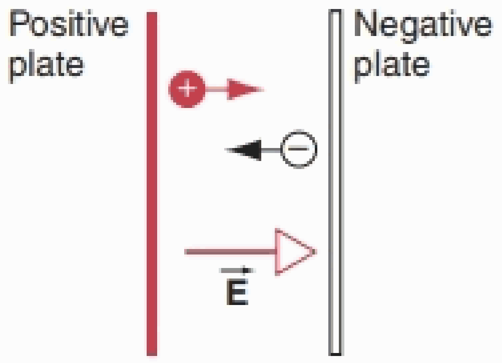


FIGURE 26-40. Problem 10.

11. An electron is projected as in Fig. 26-41 at a speed of $v_0 = 5.83 \times 10^6$ m/s and at an angle of $\theta = 39.0°$; $E = 1870$ N/C (directed upward), $d = 1.97$ cm, and $L = 6.20$ cm. Will the electron strike either of the plates? If it strikes a plate, which plate does it strike and at what distance from the left edge?

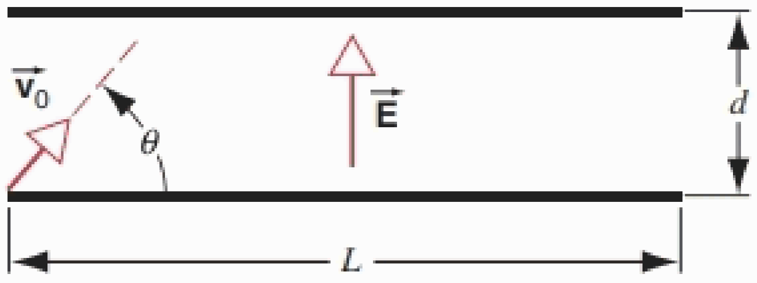


FIGURE 26-41. Problem 11.

12. An electron is constrained to move along the axis of the ring of charge discussed in Section 26-4. Show that the electron can perform small oscillations, through the center of the ring, with a frequency given by

$$\omega = \sqrt{\frac{eq}{4\pi\epsilon_0 mR^3}}.$$

13. Find the work required to turn an electric dipole end for end in a uniform electric field $\vec{\mathbf{E}}$, in terms of the dipole moment $\vec{\mathbf{p}}$ and the initial angle θ_0 between $\vec{\mathbf{p}}$ and $\vec{\mathbf{E}}$.

14. Find the frequency of oscillation of an electric dipole, of moment p and rotational inertia I, for small amplitudes of oscillation about its equilibrium position in a uniform electric field E.

15. Two equal positive point charges $+q$ are located at $z = +a/2$ and $z = -a/2$. (*a*) Derive an expression for dE_z/dz for points along the z axis, and evaluate dE_z/dz in the limit $z \ll a/2$. (*b*) Show that the force on a small dipole placed at this point, its axis along the line joining the charges, is given by $F = p(dE_z/dz)$, where p is the dipole moment and dE_z/dz is the limiting value found in part (*a*).

COMPUTER PROBLEMS

1. A ring of radius $r = 1.0$ m has a nonuniform charge density given by $\lambda = (2.0\ \mu\text{C/m})(2 + \sin\theta)$. Numerically find the coordinates of a point where the electric field vanishes.

2. The charge density on a rod of length L centered on the x axis is given by $\lambda = (1.0\ \mu\text{C/m})\sin^2(\pi x/L)$. Numerically generate a plot of the electric field lines in the xy plane.

3. Consider two particles that exert electric forces on each other. Each accelerates in response to the electric field of the other, and as their positions change the forces they exert also change. Two identical particles, each with charge $q = +1.9 \times 10^{-9}$ C and mass $m = 6.1 \times 10^{-15}$ kg, start with identical velocities of 3.0×10^4 m/s in the positive x direction. Initially one is at $x = 0$, $y = 6.7 \times 10^{-3}$ m and the other is at $x = 0$, $y = -6.7 \times 10^{-3}$ m. Both are in the xy plane and continue to move in that plane. Consider only the electric forces they exert on each other. (*a*) Use a computer program to plot the trajectories from time $t = 0$ to $t = 1.0 \times 10^{-6}$ s. Because the situation is symmetric you need calculate only the position and velocity of one of the charges. Use symmetry to find the position and velocity of the other at the beginning of each integration interval. Use $\Delta t = 1 \times 10^{-8}$ s for the integration interval. (*b*) Now suppose that one of the particles has charge $q = -1.9 \times 10^{-9}$ C, but all other conditions are the same. Plot the trajectories from $t = 0$ to $t = 5.0 \times 10^{-7}$ s.

CHAPTER 27

GAUSS' LAW

Coulomb's law can always be used to calculate the electric field $\vec{\mathbf{E}}$ *for any discrete or continuous distribution of charges at rest. The sums or integrals might be complicated (and a computer might be needed to evaluate them numerically), but the resulting electric field can always be found.*

Some cases discussed in the previous chapter used simplifying arguments based on the symmetry of the physical situation. For example, in calculating the electric field at points on the axis of a charged circular loop, we used a symmetry argument to deduce that components of $\vec{\mathbf{E}}$ *perpendicular to the axis must vanish. In this chapter we discuss an alternative to Coulomb's law, called* Gauss' law, *that provides a more useful and instructive approach to calculating the electric field in situations having certain symmetries.*

The number of situations that can directly be analyzed using Gauss' law is small, but those cases can be done with extraordinary ease. Although Gauss' law and Coulomb's law give identical results in the cases in which both can be used, Gauss' law is considered a more fundamental equation than Coulomb's law. It is fair to say that whereas Coulomb's law provides the workhorse of electrostatics, Gauss' law provides the insight.

27-1 WHAT IS GAUSS' LAW ALL ABOUT?

So far, everything we have done in electrostatics has been based on Coulomb's law, Eq. 25-4, which gives the electric force between point charges. Starting with Coulomb's law, which is essentially a mathematical representation of an experimental observation, we *defined* the electric field of a point charge q so that $\vec{\mathbf{E}} = \vec{\mathbf{F}}/q_0$, where $\vec{\mathbf{F}}$ is the force exerted on q_0 by q. By generalizing to charge distributions that can be considered as assemblies of many infinitesimal point charges, we were able to find the electric field of several different charge distributions, such as a line or a disk.

Gauss' law provides us with another way to calculate electric fields. It is equivalent to Coulomb's law for point charges, which means that everything we have done so far using Coulomb's law could also be done if we had started with Gauss' law instead.

Why do we need Gauss' law, if Coulomb's law is sufficient to calculate electric fields for any static arrangements of charges? One answer is that Gauss' law offers a much simpler way to calculate electric fields in situations with a high degree of symmetry, such as a spherical charge distribution. Another answer is that by writing Gauss' law rather than Coulomb's law as the fundamental law for electrostatics we can develop a system of equations for all electromagnetic phenomena that illustrate more clearly the relationship between electric and magnetic fields. A third reason is that Gauss' law is valid in the case of rapidly moving charges, but Coulomb's law can be used only for charges that are at rest or moving very slowly. Finally, as we show later in this chapter, Coulomb's law can be derived as a special case of Gauss' law, and thus Gauss' law is more general than Coulomb's law. For these reasons, Gauss' law is considered to be more fundamental than Coulomb's law and is included as one of the four basic equations of electromagnetism (Maxwell's equations, which we discuss in Chapter 38).

Before we introduce Gauss' law, we first need to define and discuss a new quantity, the *flux of the electric field.* The flux is a mathematical property of any field represented by vectors that is determined by the *surface integral* of the field vector over a particular area. There is also a geometrical interpretation of the flux that is based on the number of field lines that pass through the area.

27-2 THE FLUX OF A VECTOR FIELD

The word "flux" comes from a Latin word meaning "to flow," and you can consider the flux of a vector field to be a measure of the flow or penetration of the field vectors through an imaginary fixed element of surface in the field. Later we consider the flux of the electric field, but for now we consider a more familiar example, the velocity field of a flowing fluid.

Imagine a stream of fluid in steady flow, in which we represent the flow by specifying the velocity vector at each point. Figure 27-1 shows a uniform flow; the velocity vectors are parallel throughout the fluid. Suppose we place into the stream a wire bent into a square loop of area A. In Fig. 27-1*a*, the loop is placed so that its plane is perpendicular to the direction of flow. We define the flux Φ of the velocity field so that its magnitude is given by

$$|\Phi| = vA, \tag{27-1}$$

where v is the magnitude of the velocity at the location of the loop. The flux has units of m^3/s and might be considered to represent the rate at which fluid passes through the loop; in terms of the field concept (and for the purpose of introducing Gauss' law), however, it is convenient to consider the flux as a measure of the *number of field lines passing through the loop.*

In Fig. 27-1*b*, the loop has been rotated so that its plane is no longer perpendicular to the direction of the velocity. Note that the number of lines of the velocity field passing through the loop is smaller in Fig. 27-1*b* than in Fig. 27-1*a*. The projected area of the square is $A \cos \theta$, and by examining Fig. 27-1*b* you should convince yourself that the number of field lines passing through the inclined loop of area A is the same as the number of field lines passing through the smaller loop of area $A \cos \theta$ perpendicular to the stream. Thus the magnitude of the flux in the situation of Fig. 27-1*b* is

$$|\Phi| = vA \cos \theta. \tag{27-2}$$

If the loop were rotated so that the fluid velocity were parallel to its surface, as in Fig. 27-1*c*, the flux would be zero, corresponding to $\theta = 90°$ in Eq. 27-2. Note that in this case no field lines pass through the loop.

Gauss' law, as we shall see, concerns the net flux through a *closed* surface. We must therefore distinguish between positive and negative flux penetrating a surface. The right side of Eq. 27-2 can be expressed in terms of the dot product between $\vec{\mathbf{v}}$ and a vector $\vec{\mathbf{A}}$ whose magnitude is the area of the surface and whose direction is perpendicular to the surface (Fig. 27-1*d*). However, since the normal to a surface can point either in the direction shown in Fig. 27-1*d* or in the reverse direction, we must have a way to specify that direction; otherwise the sign of Φ will not be clearly defined. By convention, we choose the direction of $\vec{\mathbf{A}}$ to be that of the *outward normal* from a closed surface. Thus flux *leaving* the volume enclosed by the surface is considered positive, and flux *entering* the volume is considered negative. With this choice, we can then write the flux for a closed surface consisting of several individual surfaces (Fig. 27-1*e*, for example) as

$$\Phi = \sum \vec{\mathbf{v}} \cdot \vec{\mathbf{A}}, \tag{27-3}$$

where $\vec{\mathbf{v}}$ is the velocity vector at the surface. The sum is carried out over all the individual surfaces that make up a

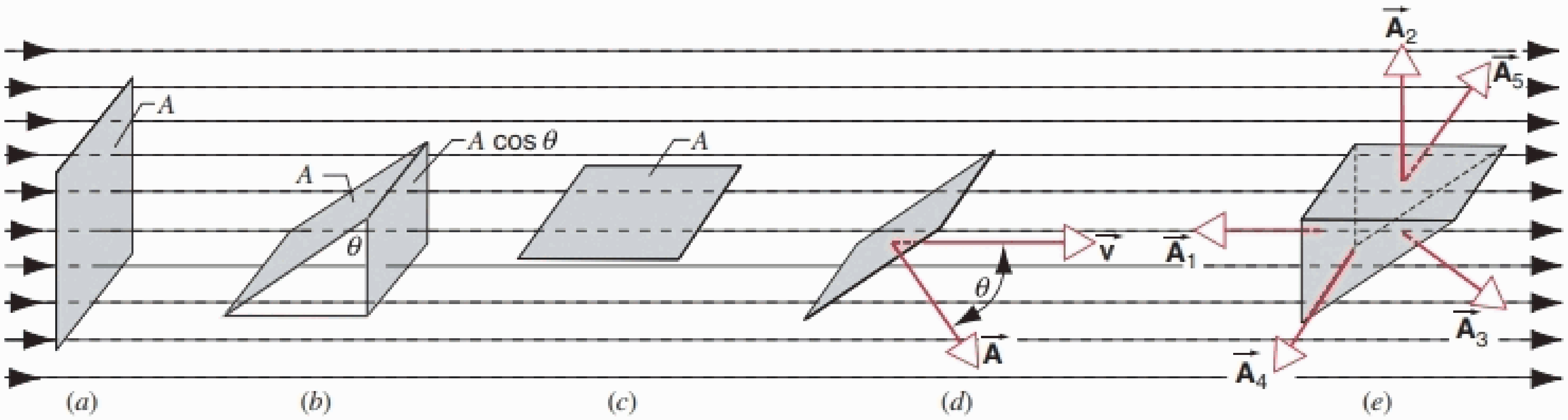


FIGURE 27-1. A wire loop of area A is immersed in a flowing stream, which we represent as a velocity field. (*a*) The loop is at right angles to the flow. (*b*) The loop is turned through an angle θ; the projection of the area perpendicular to the flow is $A \cos \theta$. (*c*) When $\theta = 90°$, none of the streamlines pass through the plane of the loop. (*d*) The area of the loop is represented by a vector $\vec{\mathbf{A}}$ perpendicular to the plane of the loop. The angle between $\vec{\mathbf{A}}$ and the flow velocity $\vec{\mathbf{v}}$ is θ. (*e*) A closed surface made of five plane surfaces. The area $\vec{\mathbf{A}}$ of each surface is represented by the outward normal.

closed surface. The flux is a scalar quantity, because it is defined in terms of the dot product of two vectors.

SAMPLE PROBLEM 27-1. Consider the closed surface of Fig. 27-1*e*, which shows a volume enclosed by five surfaces (1, 2, and 3, which are parallel to the surfaces of Figs. 27-1*a*, 27-1*c*, 27-1*b*, respectively, along with 4 and 5, which are parallel to the streamlines). Assuming the velocity field is uniform, so that it has the same magnitude and direction everywhere, find the total flux through the closed surface.

Solution Using Eq. 27-3 we can write the total flux as the sum of the values of the flux through each of the five separate surfaces:

$$\Phi = \vec{\mathbf{v}}\cdot\vec{\mathbf{A}}_1 + \vec{\mathbf{v}}\cdot\vec{\mathbf{A}}_2 + \vec{\mathbf{v}}\cdot\vec{\mathbf{A}}_3 + \vec{\mathbf{v}}\cdot\vec{\mathbf{A}}_4 + \vec{\mathbf{v}}\cdot\vec{\mathbf{A}}_5.$$

Note that for surface 1 the angle between the *outward normal* $\vec{\mathbf{A}}_1$ and the velocity $\vec{\mathbf{v}}$ is 180°, so that the dot product $\vec{\mathbf{v}}\cdot\vec{\mathbf{A}}_1$ can be written $-vA_1$. The contributions from surfaces 2, 4, and 5 all vanish, because in each case (as shown in Fig. 27-1*e*) the vector $\vec{\mathbf{A}}$ is perpendicular to $\vec{\mathbf{v}}$. For surface A_3, the flux can be written $vA_3\cos\theta$, and thus the total flux is

$$\Phi = -vA_1 + 0 + vA_3\cos\theta + 0 + 0 = -vA_1 + vA_3\cos\theta.$$

However, from the geometry of Fig. 27-1*e* we conclude that $A_3\cos\theta = A_1$, and as a result we obtain

$$\Phi = 0.$$

That is, the total flux through the closed surface is zero.

The result of the previous sample problem should not be surprising if we remember that the velocity field is an equivalent way of representing the actual flow of material particles in the stream. Every field line that enters the closed surface of Fig. 27-1*e* through surface 1 leaves through surface 3. Equivalently, we can state that, for the closed surface shown in Fig. 27-1*e*, the net amount of fluid entering the volume enclosed by the surface is equal to the net amount of fluid leaving the volume. This is to be expected *for any closed surface* if there are within the volume no *sources* or *sinks* of fluid—that is, locations at which new fluid is created or flowing fluid is trapped. If there were a source within the volume (such as a melting ice cube that introduced additional fluid into the stream), then more fluid would leave the surface than entered it, and the total flux would be positive. If there were a sink within the volume, then more fluid would enter than would leave, and the net flux would be negative. The net positive or negative flux through the surface depends on the strength of the source or sink (that is, on the volume rate at which fluid leaves the source or enters the sink). For example, if a melting solid inside the surface released 1 cm^3 of fluid per second into the stream, then we would find the net flux through the closed surface to be $+1\ \text{cm}^3/\text{s}$.

Figure 27-1 showed the special case of a uniform field and planar surfaces. We can easily generalize these concepts to a nonuniform field and to surfaces of arbitrary shape and orientation. Any arbitrary surface can be divided into infinitesimal elements of area dA that are approximately plane surfaces. The direction of the vector $d\vec{\mathbf{A}}$ is that of the outward normal to this infinitesimal element. The field has a value $\vec{\mathbf{v}}$ at the site of this element, and the net flux is found by adding the contributions of all such elements—that is, by integrating over the entire surface.

$$\Phi = \int \vec{\mathbf{v}}\cdot d\vec{\mathbf{A}}. \qquad (27\text{-}4)$$

The conclusions we derived above remain valid in this general case: if Eq. 27-4 is evaluated over a closed surface, then the flux is (1) *zero* if the surface encloses no sources or sinks, (2) *positive* and equal in magnitude to their strength if the surface contains only sources, or (3) *negative* and equal in magnitude to their strength if the surface contains only sinks. If the surface encloses both sources *and* sinks, the net flux can be zero, positive, or negative, depending on the relative strength of the sources and sinks.

In the next section we apply similar considerations to the flux of another vector field, namely, the electric field $\vec{\mathbf{E}}$. As you might anticipate, when we discuss electrostatics the sources or sinks of the field are positive or negative charges, and the strengths of the sources or sinks are proportional to the magnitudes of the charges. Gauss' law relates the flux of the electric field through a closed surface, calculated by analogy with Eq. 27-4, to the net electric charge enclosed by the surface.

27-3 THE FLUX OF THE ELECTRIC FIELD

Imagine the field lines in Fig. 27-1 to represent an electric field of charges at rest rather than a velocity field. Even though nothing is flowing in the electrostatic case, we still use the concept of flux. The definition of electric flux is similar to that of velocity flux, with $\vec{\mathbf{E}}$ replacing $\vec{\mathbf{v}}$ wherever it appears. In analogy with Eq. 27-3, we define the flux of the electric field Φ_E as

$$\Phi_E = \sum \vec{\mathbf{E}}\cdot\vec{\mathbf{A}}. \qquad (27\text{-}5)$$

As was the case with the velocity flux, the flux Φ_E can be considered as a measure of the number of lines of the electric field that passes through the surface. The subscript E on Φ_E reminds us that we are speaking of the *electric* flux and serves to distinguish electric from magnetic flux, which we consider in Chapter 34. Equation 27-5 applies, as did Eq. 27-3, only to cases in which $\vec{\mathbf{E}}$ is constant in magnitude and direction over each area $\vec{\mathbf{A}}$ included in the sum.

Like the velocity flux, the flux of the electric field is a scalar. Its units are, from Eq. 27-5, $\text{N}\cdot\text{m}^2/\text{C}$.

Gauss' law deals with the flux of the electric field through a closed surface. To define Φ_E more generally, particularly in cases in which $\vec{\mathbf{E}}$ is not uniform, consider

Fig. 27-2, which shows an arbitrary closed surface immersed in a nonuniform electric field. Let us divide the surface into small squares of area ΔA, each of which is small enough so that it may be considered to be plane. Each element of area can be represented as a vector $\Delta\vec{\mathbf{A}}$ whose magnitude is the area ΔA. The direction of $\Delta\vec{\mathbf{A}}$ is taken as the outward-drawn normal to the surface, as in Fig. 27-1. Since the squares have been made very small, $\vec{\mathbf{E}}$ may be taken as constant for all points on a given square.

The vectors $\vec{\mathbf{E}}$ and $\Delta\vec{\mathbf{A}}$ that characterize each square make an angle θ with each other. Figure 27-2 shows an enlarged view of three squares on the surface, marked *a*, *b*, and *c*. Note that at *a*, $\theta > 90°$ ($\vec{\mathbf{E}}$ points in); at *b*, $\theta = 90°$ ($\vec{\mathbf{E}}$ is parallel to the surface); and at *c*, $\theta < 90°$ ($\vec{\mathbf{E}}$ points out).

A provisional definition of the total flux of the electric field over the surface is, by analogy with Eq. 27-5,

$$\Phi_E = \sum \vec{\mathbf{E}} \cdot \Delta\vec{\mathbf{A}}, \qquad (27\text{-}6)$$

which instructs us to add up the scalar quantity $\vec{\mathbf{E}} \cdot \Delta\vec{\mathbf{A}}$ for all elements of area into which the surface has been divided. For points such as *a* in Fig. 27-2 the contribution to the flux is negative; at *b* it is zero, and at *c* it is positive. Thus if $\vec{\mathbf{E}}$ is everywhere outward ($\theta < 90°$), each $\vec{\mathbf{E}} \cdot \Delta\vec{\mathbf{A}}$ is positive, and Φ_E for the entire surface is positive. If $\vec{\mathbf{E}}$ is everywhere inward ($\theta > 90°$), each $\vec{\mathbf{E}} \cdot \Delta\vec{\mathbf{A}}$ is negative, and Φ_E for the surface is negative. Whenever $\vec{\mathbf{E}}$ is everywhere parallel to a surface ($\theta = 90°$), each $\vec{\mathbf{E}} \cdot \Delta\vec{\mathbf{A}}$ is zero, and Φ_E for the surface is zero.

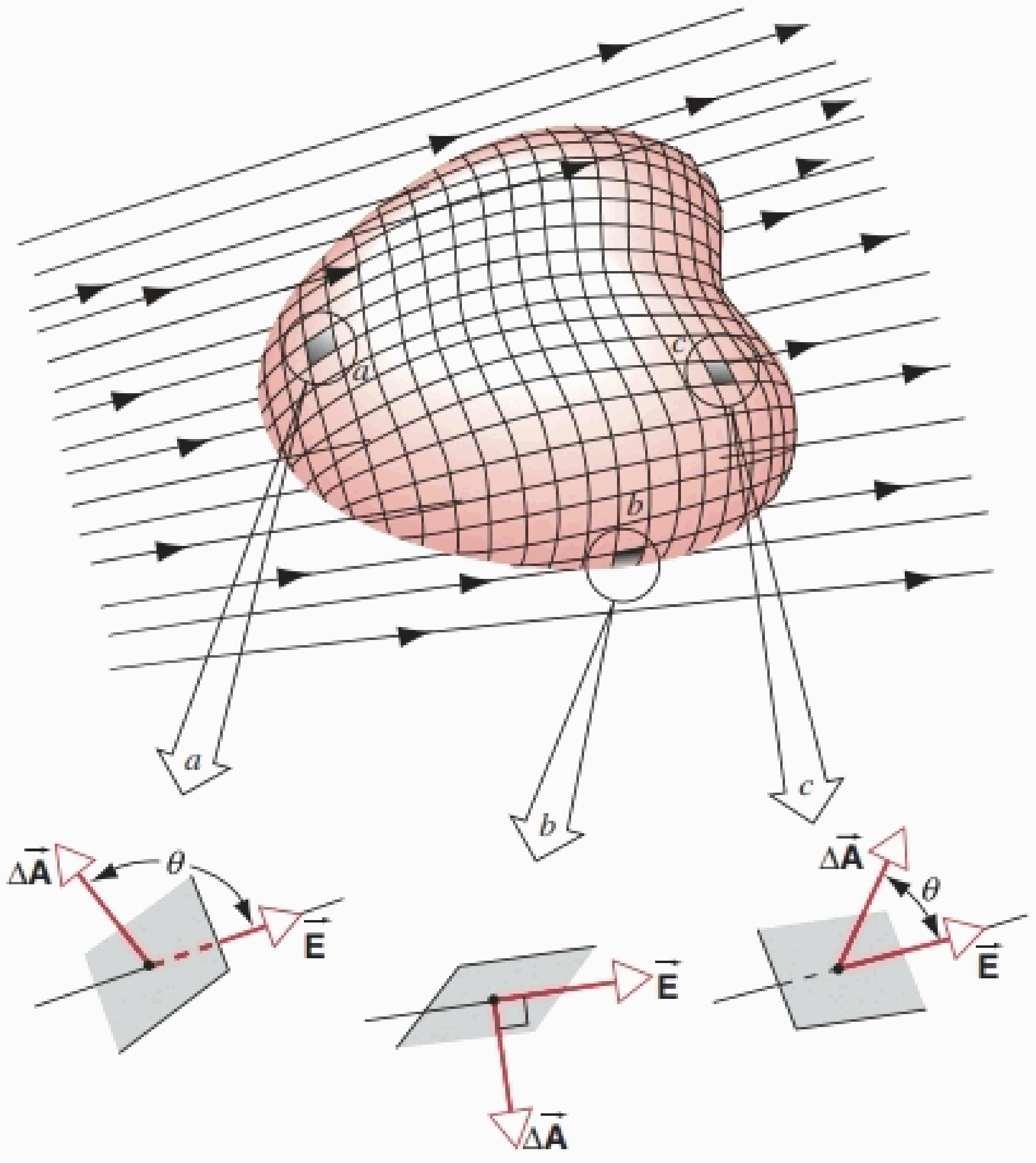


FIGURE 27-2. A surface of arbitrary shape immersed in a nonuniform electric field $\vec{\mathbf{E}}$. The surface is divided into small elements of area $\Delta\vec{\mathbf{A}}$. The relationship between the vectors $\vec{\mathbf{E}}$ and $\Delta\vec{\mathbf{A}}$ is shown for three different elements (*a*, *b*, and *c*).

The exact definition of electric flux is found in the differential limit of Eq. 27-6. Replacing the sum over the surface by an integral over the surface yields

$$\Phi_E = \int \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}. \qquad (27\text{-}7)$$

This *surface integral* indicates that the surface in question is to be divided into infinitesimal elements of area $d\vec{\mathbf{A}}$ and that the scalar quantity $\vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$ is to be evaluated for each element and summed over the entire surface. In the case of Gauss' law, we are concerned with evaluating this integral over a *closed* surface. In this case the integral sign is written with a circle $\oint$ as a reminder.

SAMPLE PROBLEM 27-2. Figure 27-3 shows a hypothetical closed cylinder of radius R immersed in a uniform electric field $\vec{\mathbf{E}}$, the cylinder axis being parallel to the field. What is Φ_E for this closed surface?

Solution The flux Φ_E can be written as the sum of three terms, an integral over (*a*) the left cylinder cap, (*b*) the cylindrical surface, and (*c*) the right cap. Thus, from Eq. 27-7, written for a closed surface,

$$\begin{aligned}\Phi_E &= \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} \\ &= \int_a \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} + \int_b \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} + \int_c \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}.\end{aligned}$$

For the left cap, the angle θ for all points is 180°, $\vec{\mathbf{E}}$ has a constant value, and the vectors $d\vec{\mathbf{A}}$ are all parallel. Thus

$$\int_a \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \int E\, dA \cos 180° = -E \int dA = -EA,$$

where A ($= \pi R^2$) is the area of the left cap. Similarly, for the right cap,

$$\int_c \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = +EA,$$

the angle θ for all points being 0 here. Finally, for the cylinder wall,

$$\int_b \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = 0,$$

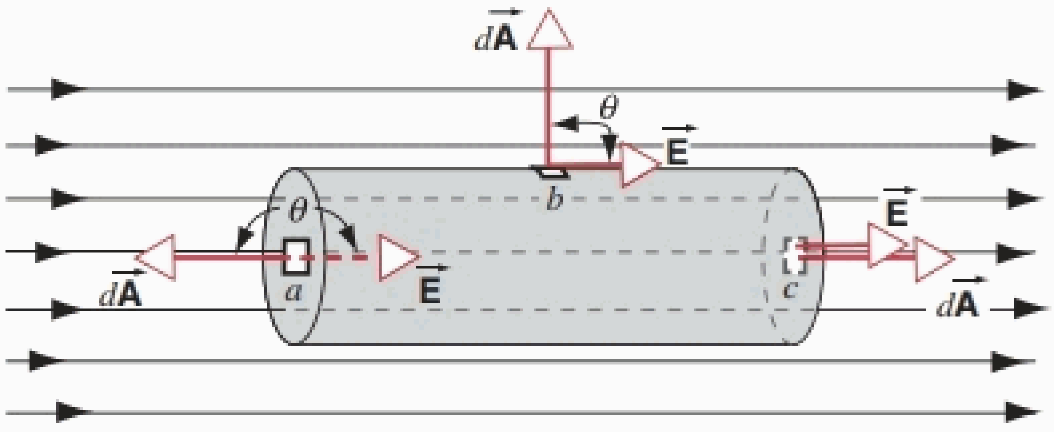


FIGURE 27-3. Sample Problem 27-2. A closed cylinder is immersed in a uniform electric field $\vec{\mathbf{E}}$ parallel to its axis.

because $\theta = 90°$; hence $\vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = 0$ for all points on the cylindrical surface. Thus the total flux is

$$\Phi_E = -EA + 0 + EA = 0.$$

This result is expected, because there are no charges within the closed surface of Fig. 27-5. Lines of (constant) $\vec{\mathbf{E}}$ enter at the left and emerge at the right, just as in Fig. 27-1*e*.

Flux and Lines of Field

To illustrate the relationship between flux and number of electric field lines passing through a closed surface, let us agree that each unit of charge q will be represented by a specific number of field lines, for example six lines, as shown in Fig. 27-4.* Six field lines point away from a charge $+q$, and six field lines terminate on a charge $-q$. If each charge is surrounded by a closed surface, then the electric flux through the surface surrounding the positive charge is $+6$ units, and the electric flux through the surface surrounding the negative charge is -6 units. (We count $+1$ arbitrary unit of flux for field lines that pass outward through the surface and -1 unit for field lines that pass inward through the surface.) No matter how big or how small is the surface surrounding each charge, six field lines always penetrate the surface and the flux is six units.

In Fig. 27-4*b*, the field line at the bottom of the drawing passes through the surface three times. Moving toward the charge from outside the surface, the first time the field line enters the surface we count -1, because it is going inward through the surface; the second time we count $+1$ because it is now going outward, and the third passage inward through the surface gives us another -1. The net contribution to the flux through the surface for this field line is -1, and the net flux for the entire surface is -6 units. No matter what the shape of the surface is or how it is stretched or distorted, the net flux through the surface is the same and is determined only by how much charge the surface encloses.

Figure 27-5 shows a closed surface enclosing charges $+q$ and $-q$. The net flux through the surface is zero, because for every field line from the positive charge passing outward through the surface, there is a field line from the negative charge passing inward through the surface. Because both charges have the same magnitude, the total number of field lines is zero, and therefore the net flux is zero.

Now suppose you are told that there are $+30$ field lines (or $+30$ units of flux) passing through an arbitrary closed surface. Can you determine how much charge is contained within the surface and where it is located inside the surface? We know that the *net* charge inside the surface is $+5q$, but we do not know whether one particle contains the full charge $+5q$, or whether there are two particles of charges $+6q$ and $-q$, three particles of $+8q$, $+4q$, and $-7q$, or any one of an infinite number of other possibilities. Also, the charge or charges can be located anywhere inside the surface and still produce the same $+30$ units of flux. If we know only the flux, we know the net amount of charge inside the surface, but we cannot deduce anything about the size or locations of the charges, and we therefore cannot deduce anything about the electric field, either on the surface or elsewhere in space.

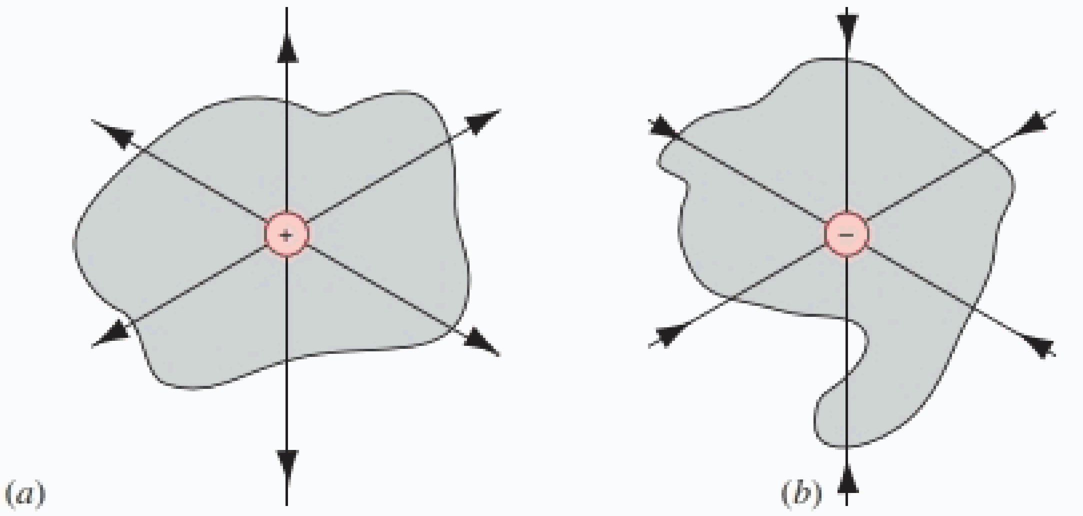


FIGURE 27-4. (*a*) Six field lines pass through an arbitrary closed surface surrounding a positive charge $+q$. (*b*) Six field lines enter the surface surrounding a negative charge $-q$.

If, however, we draw a spherical surface, and if we know that the flux is distributed uniformly over the surface, then we can conclude that all of the charge is located at the center of the sphere in a single particle of charge $+5q$, and knowing the magnitude and location of this charged particle we can then deduce the electric field at any location. We therefore conclude that

> *The relationship between the total flux through a closed surface and the net charge enclosed by the surface is always valid, but we can use this relationship to deduce*

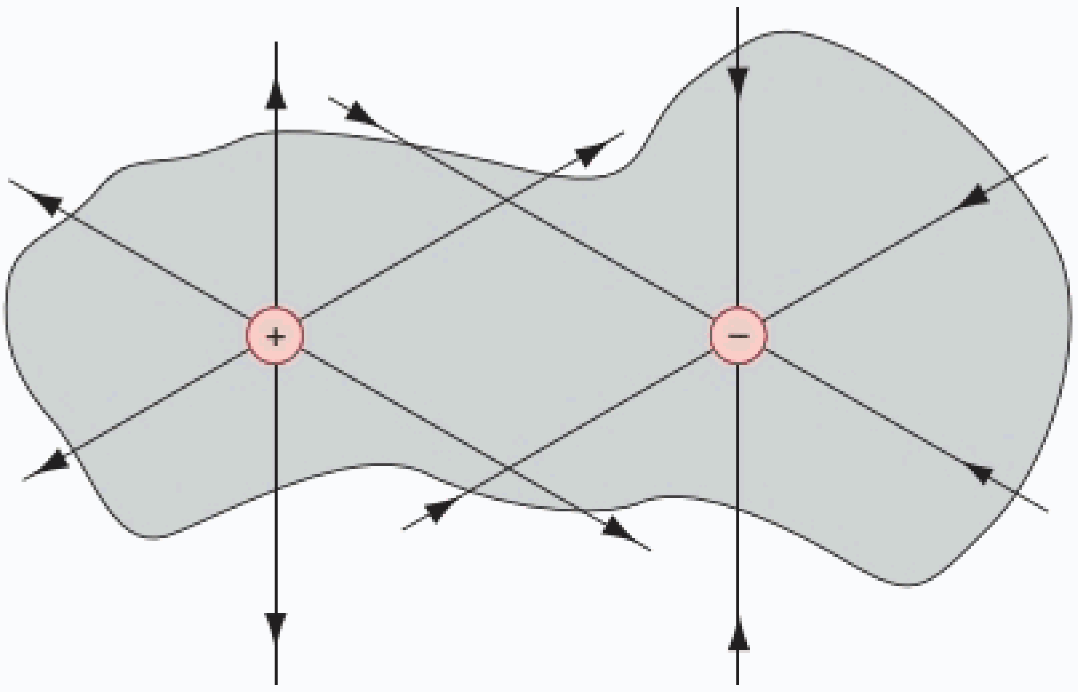


FIGURE 27-5. If the net charge enclosed by a surface is zero, then the total number of field lines (and the total electric flux) passing through the surface is zero.

---

*For convenience, we represent the field line diagrams in two dimensions rather than three. In three dimensions, the diagrams are often more complex and care must be taken to represent the field pattern associated with the charges. For a discussion see "Electric Field Line Diagrams Don't Work," by A. Wolf, S. J. Van Hook, and E. R. Weeks, *American Journal of Physics,* June 1996, p. 714.

the electric field at points in space only if the geometry of the charge and the surface has a high degree of symmetry.

Gauss' law gives the relationship between the flux through a closed surface and the net charge enclosed by the surface.

27-4 GAUSS' LAW*

Now that we have defined the flux of the electric field vector through a *closed* surface, we are ready to write Gauss' law. Let us suppose we have a collection of positive and negative charges, which establish an electric field $\vec{\mathbf{E}}$ throughout a certain region of space. We construct in that space an imaginary closed surface called a *Gaussian surface,* which may or may not enclose some of the charges. Gauss' law, which relates the total flux Φ_E through this surface to the *net* charge q enclosed by the surface, can be stated as

$$\epsilon_0 \Phi_E = q \tag{27-8}$$

or

$$\epsilon_0 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q. \tag{27-9}$$

The circle on the integral sign indicates that the integral is to be carried out over a *closed* surface. We see that Gauss' law predicts that Φ_E is zero for the surface considered in Sample Problem 27-2, because the surface encloses no charge.

As discussed in Section 26-5, the magnitude of the electric field is proportional to the number of field lines crossing an element of area perpendicular to the field. The integral in Eq. 27-9 essentially counts the number of field lines passing through the surface. It is entirely reasonable that the number of field lines passing through a surface should be proportional to the net charge enclosed by the surface, as Eq. 27-9 requires.

The choice of the Gaussian surface is arbitrary. It is usually chosen so that the symmetry of the distribution gives, on at least part of the surface, an electric field of constant magnitude, which can then be factored out of the integral of Eq. 27-9. In such a situation, Gauss' law can be used to evaluate the electric field.

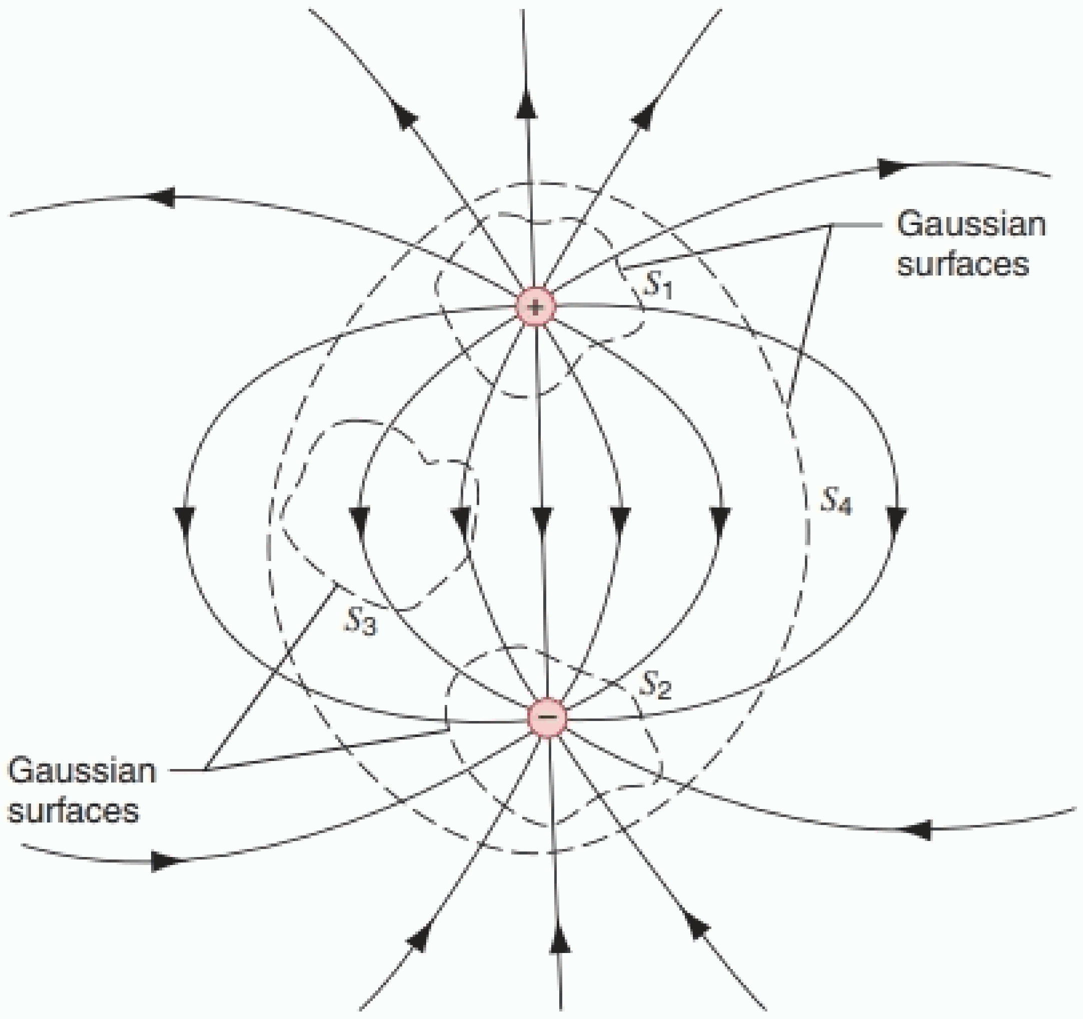


FIGURE 27-6. Two equal and opposite charges and the lines that represent the electric field in their vicinity. The cross sections of four closed Gaussian surfaces are shown.

Figure 27-6 shows the lines of force (and thus of electric field) of a dipole. Four closed Gaussian surfaces have been drawn, the cross sections of which are shown in the figure. On surface S_1, the electric field is everywhere outward from the surface and thus, as was the case with surface element c of Fig. 27-2, $\vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$ is everywhere positive on S_1. When we evaluate the integral of Eq. 27-9 over the entire closed surface, we get a positive result. Equation 27-9 then demands that the surface must enclose a net positive charge, as is the case. In Faraday's terminology, more lines of force leave the surface than enter it, so it must enclose a net positive charge.

On surface S_2 of Fig. 27-6, on the other hand, the electric field is everywhere entering the surface. Like surface element a in Fig. 27-2, $\vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$ is negative for every element of area, and the integral of Eq. 27-9 gives a negative value, which indicates that the surface encloses a net negative charge (as is the case). More lines of force enter the surface than leave it.

Surface S_3 encloses no charge at all, so according to Gauss' law the total flux through the surface must be zero. This is consistent with Fig. 27-6, which shows that as many lines of force enter the top of the surface as leave the bottom. This is no accident; you can draw a surface in Fig. 27-6 of any irregular shape, and as long as it encloses neither of the charges, the number of field lines that enter the surface equals the number that leave the surface.

Surface S_4 also encloses no *net* charge, since we assumed the magnitudes of the two charges to be equal. Once again, the total flux through the surface should be zero. Some of the field lines are wholly contained within the surface and therefore do not contribute to the flux *through* the surface. However, since every field line that leaves the positive charge eventually terminates on the negative charge, every line from the positive charge that breaks the surface in the outward direction has a corresponding line that breaks the surface in the inward direction as it seeks the negative charge. The total flux is therefore zero.

---

*Karl Friedrich Gauss (1777–1855) was a German mathematician who made substantial discoveries in number theory, geometry, and probability. He also contributed to astronomy and to measuring the size and shape of the Earth. See "Gauss," by Ian Stewart, *Scientific American,* July 1977, p. 122, for a fascinating account of the life of this remarkable mathematician.

Gauss' Law and Coulomb's Law

Coulomb's law can be deduced from Gauss' law and symmetry considerations. To do so, let us apply Gauss' law to an isolated positive point charge q as in Fig. 27-7. Although Gauss' law holds for any surface whatever, we choose a spherical surface of radius r centered on the charge. The advantage of this surface is that, from symmetry, $\vec{\mathbf{E}}$ must be perpendicular to the surface, so the angle θ between $\vec{\mathbf{E}}$ and $d\vec{\mathbf{A}}$ is zero everywhere on the surface. Moreover, $\vec{\mathbf{E}}$ has the same magnitude everywhere on the surface. *Constructing a Gaussian surface that takes advantage of such a symmetry is of fundamental importance in applying Gauss' law.*

In Fig. 27-7 both $\vec{\mathbf{E}}$ and $d\vec{\mathbf{A}}$ at any point on the Gaussian surface are directed radially outward, so the quantity $\vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$ becomes simply $E\,dA$. Gauss' law (Eq. 27-9) thus reduces to

$$\epsilon_0 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \epsilon_0 \oint E\,dA = q.$$

Because E has the same magnitude for all points on the sphere, it can be factored from inside the integral sign, which gives

$$\epsilon_0 E \oint dA = q.$$

The integral is simply the total surface area of the sphere, $4\pi r^2$. We therefore obtain

$$\epsilon_0 E(4\pi r^2) = q$$

or

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}. \qquad (27\text{-}10)$$

Equation 27-10 gives the magnitude of the electric field $\vec{\mathbf{E}}$ at any point a distance r from an isolated point charge q and is identical to Eq. 26-6, which was obtained from Coulomb's law. Thus by choosing a Gaussian surface with the proper symmetry, we obtain Coulomb's law from Gauss' law. These two laws can be regarded as equivalent for our applications, but (as we discussed in Section 27-1) Gauss' law is more generally applicable and so is regarded as a more fundamental equation of electromagnetism.

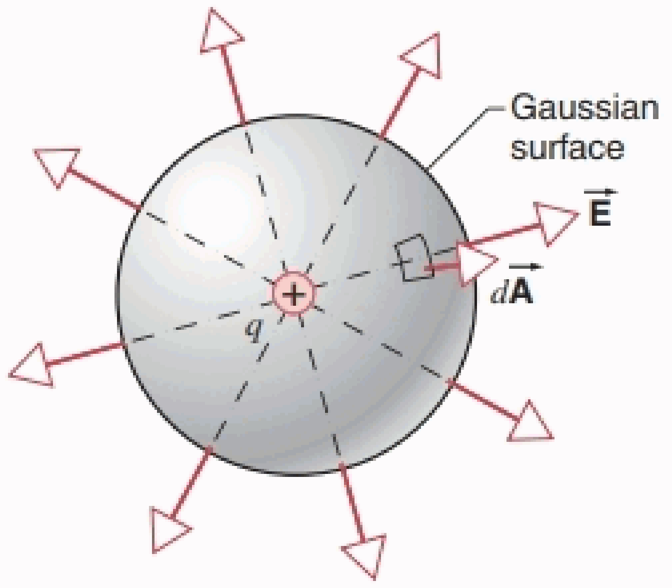


FIGURE 27-7. A spherical Gaussian surface of radius r surrounding a positive point charge q.

It is interesting to note that writing the proportionality constant in Coulomb's law as $1/4\pi\epsilon_0$ permits a simpler form for Gauss' law. If we had written the Coulomb law constant simply as K, Gauss' law would have to be written as $(1/4\pi K)\Phi_E = q$. We prefer to leave the factor 4π in Coulomb's law so that it will not appear in Gauss' law or in other frequently used relations that are derived later.

27-5 APPLICATIONS OF GAUSS' LAW

Gauss' law can be used to calculate $\vec{\mathbf{E}}$ if the symmetry of the charge distribution is high. One example of this calculation, the field of a point charge, has already been discussed in connection with Eq. 27-10. Here we present other examples.

Infinite Line of Charge

Figure 27-8 shows a section of an infinite line of charge of constant positive linear charge density (charge per unit length) λ. We would like to find the electric field at a distance r from the line.

In Section 26-4 we discussed the symmetry arguments that lead us to conclude that the electric field in this case can have only a radial component. The problem therefore has cylindrical symmetry, and so as a Gaussian surface we choose a circular cylinder of radius r and length h, closed at each end by plane caps normal to the axis. E is constant over the cylindrical surface and perpendicular to the surface. The flux of $\vec{\mathbf{E}}$ through this surface is $E(2\pi rh)$, where $2\pi rh$ is the area of the surface. There is no flux through the circular caps because $\vec{\mathbf{E}}$ here is parallel to the surface at every point, so that $\vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = 0$ everywhere on the caps.

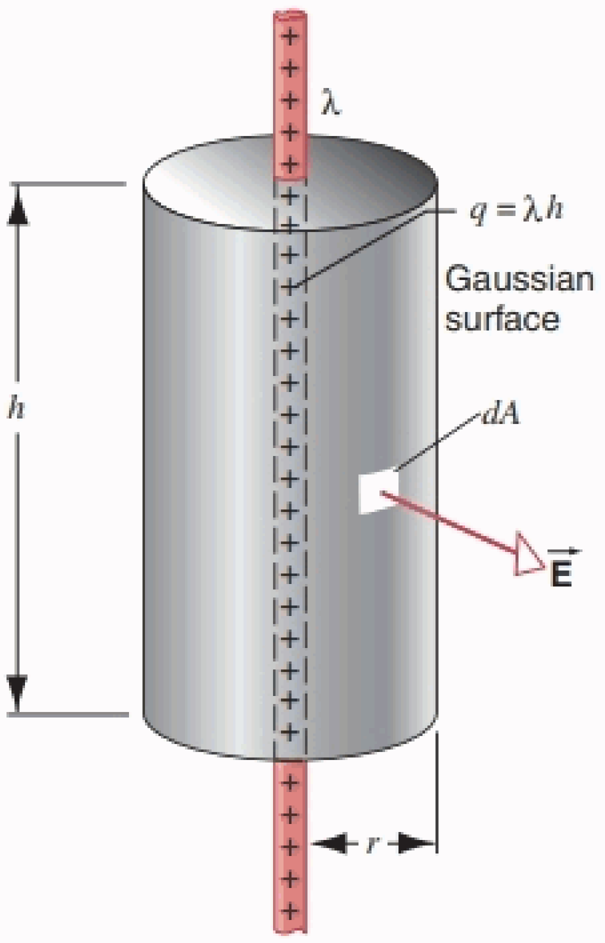


FIGURE 27-8. A Gaussian surface in the shape of a closed cylinder surrounds a portion of an infinite line of positive charge.

The charge q enclosed by the Gaussian surface of Fig. 27-8 is λh. Gauss' law (Eq. 27-9) then gives

$$\epsilon_0 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q$$

$$\epsilon_0 E(2\pi r h) = \lambda h,$$

or

$$E = \frac{\lambda}{2\pi\epsilon_0 r}, \tag{27-11}$$

in agreement with Eq. 26-17.

Note how much simpler is the solution using Gauss' law than that using integration methods, as in Chapter 26. Note too that the solution using Gauss' law is possible only if we choose our Gaussian surface to take full advantage of the cylindrical symmetry of the electric field set up by a long line of charge. We are free to choose any closed surface, such as a cube or a sphere (see Exercise 24), for a Gaussian surface. Even though Gauss' law holds for all such surfaces, they are not all useful for the problem at hand; only the cylindrical surface of Fig. 27-8 is appropriate in this case.

Gauss' law has the property that it provides a useful technique for calculation only in problems that have a certain degree of symmetry, but in these problems the solutions are strikingly simple.

Infinite Sheet of Charge

Figure 27-9 shows a portion of a thin, nonconducting, infinite sheet of charge of constant positive surface charge density σ (charge per unit area). We calculate the electric field at points near the sheet.

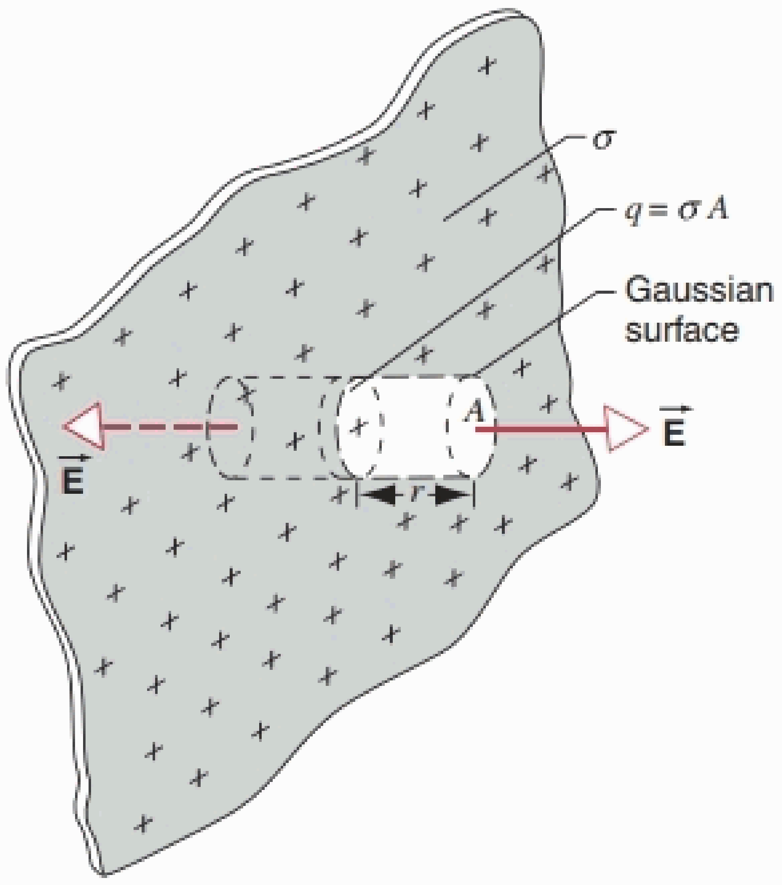


FIGURE 27-9. A Gaussian surface in the form of a small closed cylinder intersects a small portion of a sheet of positive charge. The field is perpendicular to the sheet, and so only the end caps of the Gaussian surface contribute to the flux.

A convenient Gaussian surface is a closed cylinder of cross-sectional area A, arranged to pierce the plane as shown. From symmetry, we can conclude that $\vec{\mathbf{E}}$ points at right angles to the end caps and away from the plane. Since $\vec{\mathbf{E}}$ does not pierce the cylindrical surface, there is no contribution to the flux from the curved wall of the cylinder. We assume that the end caps are equidistant from the sheet. From symmetry the field has the same magnitude at the end caps. The flux through each end cap is EA and is positive for both. Gauss' law gives

$$\epsilon_0 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q$$

$$\epsilon_0(EA + EA) = \sigma A,$$

where σA is the enclosed charge. Solving for E, we obtain

$$E = \frac{\sigma}{2\epsilon_0}. \tag{27-12}$$

Note that E is the same for all points on each side of the sheet.

Although an infinite sheet of charge cannot exist physically, this derivation is still useful in that Eq. 27-12 gives approximately correct results for real (not infinite) charge sheets if we consider only points that are far from the edges and whose distance from the sheet is small compared to the dimensions of the sheet. In fact, Eq. 27-12 agrees with Eq. 26-20, which we obtained by considering points close to a circular disk of charge.

A Spherical Shell of Charge

In Section 25-5 we used the similarity between electrostatic and gravitational forces to establish two properties of the forces exerted by uniformly charged spherical shells. Then in Section 26-4 we used those properties of the electrostatic force to deduce the electric field due to a uniformly charged spherical shell at points inside or outside the shell.

We can summarize the *shell theorems* for electric fields as follows:

1. *A uniform spherical shell of charge behaves, for external points, as if all its charge were concentrated at its center.*
2. *A uniform spherical shell of charge exerts no electrical force on a charged particle placed inside the shell.*

Let us see how Gauss' law simplifies the calculation of the electric field and the proofs of these shell theorems in this very symmetric geometry. Figure 27-10 shows a thin spherical shell on which a charge q is uniformly distributed. The shell is surrounded by two concentric spherical Gaussian surfaces, S_1 and S_2. From a symmetry argument, we conclude that the field can have only a radial component E_r. (Assume that there were a nonradial component, and suppose that someone rotated the shell

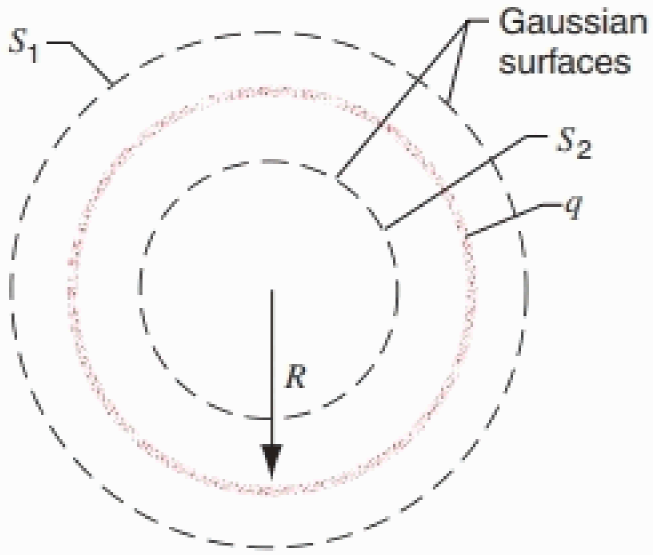


FIGURE 27-10. A cross section of a thin uniformly charged shell of total charge q. The shell is surrounded by two closed spherical Gaussian surfaces, one inside the shell and another outside the shell.

through some angle about a diameter when your back was turned. When you turned back, you could use a probe of the electric field—say, a test charge—to learn that the electric field had changed direction, even though the charge distribution was the same as before the rotation. Clearly this is a contradiction. Would this symmetry argument hold if the charge were *not* uniformly distributed over the surface?) Applying Gauss' law to surface S_1, for which $r > R$, gives

$$\epsilon_0 E_r(4\pi r^2) = q,$$

or

$$E_r = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} \qquad \text{(spherical shell, } r > R\text{)}, \qquad (27\text{-}13)$$

just as it did in connection with Fig. 27-7. Thus *the uniformly charged shell behaves like a point charge for all points outside the shell.* This proves the first shell theorem.

Applying Gauss' law to surface S_2, for which $r < R$, leads directly to

$$E_r = 0 \qquad \text{(spherical shell, } r < R\text{)}, \qquad (27\text{-}14)$$

because this Gaussian surface encloses no charge and because E_r (by another symmetry argument) has the same value everywhere on the surface. *The electric field therefore vanishes inside a uniform shell of charge;* a test charge placed anywhere in the interior would feel no electric force. This proves the second shell theorem.

These two theorems apply only in the case of a *uniformly* charged shell. If the charges were sprayed on the surface in a nonuniform manner, such that the charge density varied over the surface, these theorems would not apply. The symmetry would be lost, and as a result $\vec{\mathbf{E}}$ could not be removed from the integral in Gauss' law. The flux would remain equal to q/ϵ_0 for all exterior surfaces and zero for all interior surfaces, but we would not be able to make such a direct connection with $\vec{\mathbf{E}}$ as we can in the uniform case. In contrast to the uniformly charged shell, the field would *not* be zero throughout the interior.

Spherically Symmetric Charge Distribution

Figure 27-11 shows a cross section of a spherical distribution of charge of radius R. Here the charge is distributed throughout the spherical volume. We do not assume that the volume charge density ρ (the charge per unit volume) is a constant; however, we make the restriction that ρ at any point depends *only* on the distance of the point from the center, a condition called *spherical symmetry.* That is, ρ may be a function of r, but not of any angular coordinate. Let us find an expression for E for points outside (Fig. 27-11*a*) and inside (Fig. 27-11*b*) the charge distribution.

Any spherically symmetric charge distribution, such as that of Fig. 27-11, can be regarded as a nest of concentric thin shells. The volume charge density ρ may vary from one shell to the next, but we make the shells so thin that we can assume ρ is uniform on any particular shell. We can use the results of the previous subsection to calculate the contribution of each shell to the total electric field. The electric field from each thin shell has only a radial component, and thus the total electric field of the sphere can likewise have only a radial component. (This conclusion also follows from a symmetry argument but would not hold if the charge distribution lacked spherical symmetry—that is, if ρ depended on direction.)

Let us calculate the radial component of the electric field at points that lie at a distance r greater than the radius R of the sphere, as shown in Fig. 27-11*a*. Each concentric shell, with a charge dq, contributes a radial component dE_r to the electric field according to Eq. 27-13. The total field is the total of all such components, and because all components to the field are radial, we must compute only an algebraic sum rather than a vector sum. The sum over all the shells then gives

$$E_r = \int dE_r = \int \frac{1}{4\pi\epsilon_0}\frac{dq}{r^2}$$

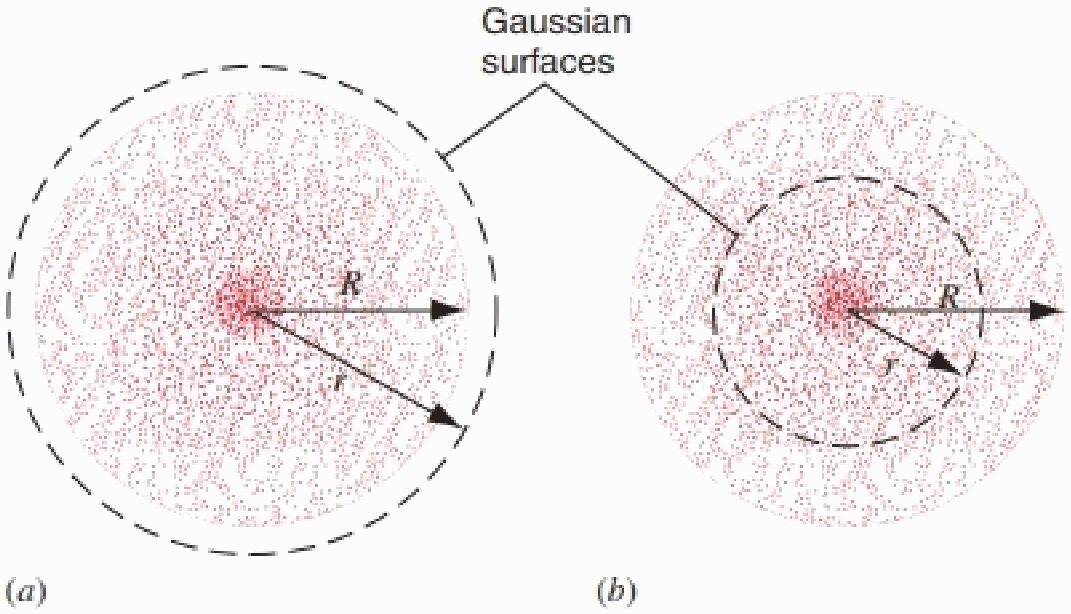


FIGURE 27-11. A cross section of a spherically symmetric charge distribution, in which the volume charge density may vary with r in this assumed nonconducting material. Closed spherical Gaussian surfaces have been drawn (*a*) outside the distribution and (*b*) within the distribution.

or, since r is constant in the integral over q,

$$E_r = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}, \tag{27-15}$$

where q is the total charge of the sphere. Thus for points outside a spherically symmetric distribution of charge, the electric field has the value that it would have if the charge were concentrated at its center. This result is similar to the gravitational case proved in Section 14-5. Both results follow from the inverse square nature of the corresponding force laws.

We now consider the electric field for points inside the charge distribution. Figure 27-11b shows a spherical Gaussian surface of radius $r < R$. Gauss' law gives

$$\epsilon_0 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \epsilon_0 E_r(4\pi r^2) = q'$$

or

$$E_r = \frac{1}{4\pi\epsilon_0}\frac{q'}{r^2}, \tag{27-16}$$

in which q' is that part of q contained *within* the sphere of radius r. According to the second shell theorem, the part of q that lies outside this sphere makes no contribution to $\vec{\mathbf{E}}$ at radius r.

To continue this calculation, we must know the charge q' that is within the radius r; that is, we must know $\rho(r)$. Let us consider the special case in which the sphere is uniformly charged, so that the charge density ρ has the same value for all points within a sphere of radius R and is zero for all points outside this sphere. For points inside such a uniform sphere of charge, the fraction of charge within r is equal to the fraction of the volume within r, and so

$$\frac{q'}{q} = \frac{\frac{4}{3}\pi r^3}{\frac{4}{3}\pi R^3}$$

or

$$q' = q\left(\frac{r}{R}\right)^3,$$

where $\frac{4}{3}\pi R^3$ is the volume of the spherical charge distribution. The expression for E_r then becomes

$$E_r = \frac{1}{4\pi\epsilon_0}\frac{qr}{R^3} \quad \text{(uniform sphere, } r < R\text{)}. \tag{27-17}$$

in agreement with Eq. 26-24. This equation becomes zero, as it should, for $r = 0$. Equation 27-17 applies *only* when the charge density is uniform, independent of r. Note that Eqs. 27-15 and 27-17 give the same result, as they must, for points on the surface of the charge distribution (that is, for $r = R$). Figure 27-12 shows the electric field for points with $r < R$ (given by Eq. 27-17) and for points with $r > R$ (given by Eq. 27-15).

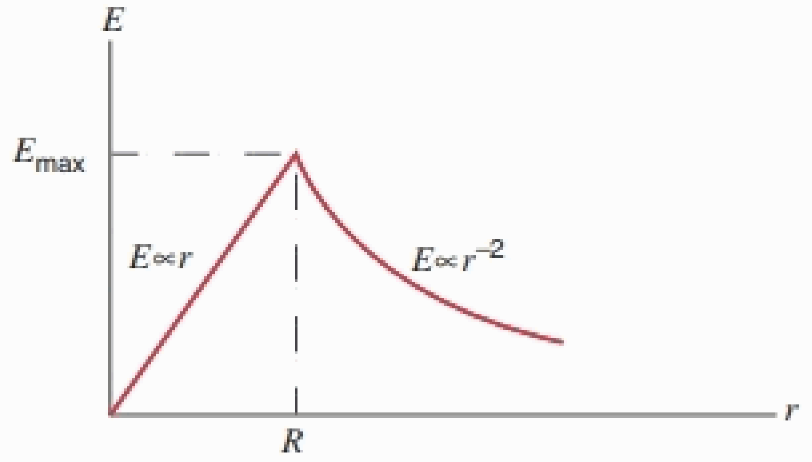


FIGURE 27-12. The variation with radius of the electric field due to a uniform spherical distribution of charge of radius R. The variation for $r > R$ applies to *any* spherically symmetric charge distribution, whereas that for $r < R$ applies *only* to a uniform distribution.

SAMPLE PROBLEM 27-3. Figure 27-13a shows portions of two large sheets of charge with uniform surface charge densities of $\sigma_+ = +6.8\ \mu\text{C/m}^2$ and $\sigma_- = -4.3\ \mu\text{C/m}^2$. Find the electric field $\vec{\mathbf{E}}$ to the left of the sheets, between the sheets, and to the right of the sheets.

Solution Our strategy is to deal with each sheet separately and then to add the resulting electric fields by using the superposition principle. For the positive sheet we have, from Eq. 27-12,

$$E_+ = \frac{\sigma_+}{2\epsilon_0} = \frac{6.8 \times 10^{-6}\ \text{C/m}^2}{(2)(8.85 \times 10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2)} = 3.84 \times 10^5\ \text{N/C}.$$

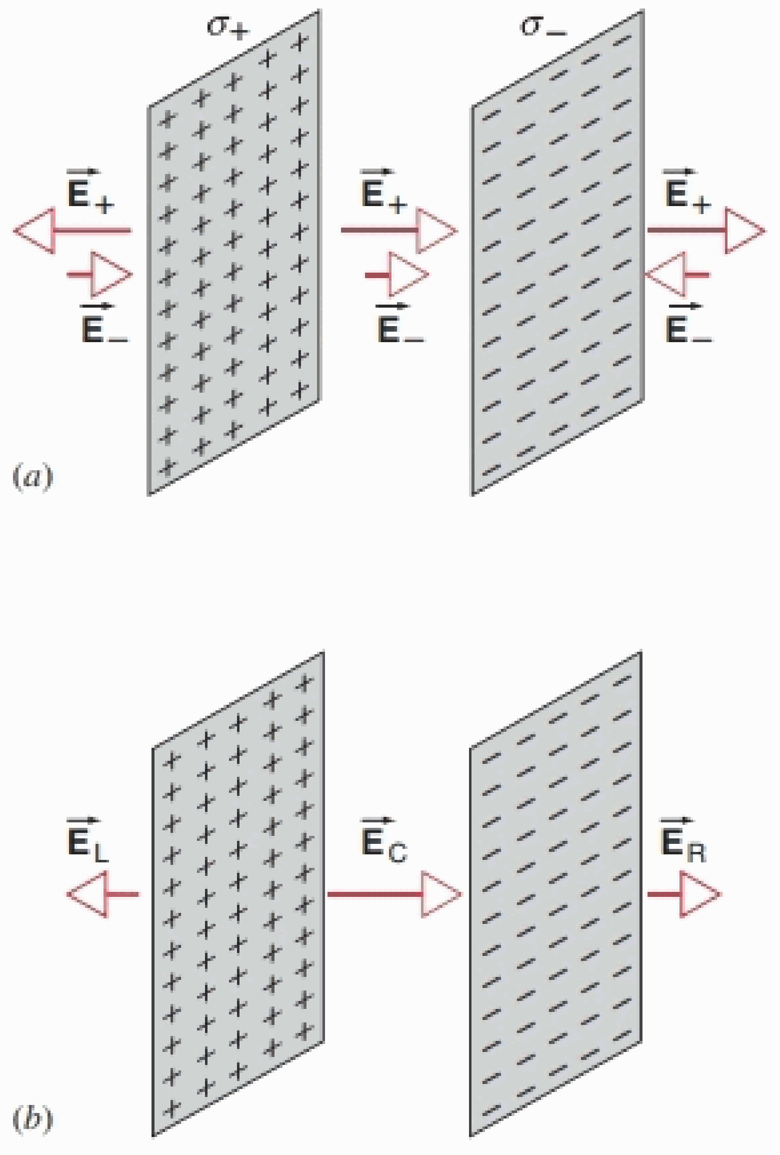


FIGURE 27-13. Sample Problem 27-3. (a) Two large parallel sheets of charge carry different charge distributions σ_+ and σ_-. The fields $\vec{\mathbf{E}}_+$ and $\vec{\mathbf{E}}_-$ would be set up by each sheet if the other were not present. (b) The net fields in the nearby regions to the left (L), center (C), and right (R) of the sheets, calculated from the vector sum of $\vec{\mathbf{E}}_+$ and $\vec{\mathbf{E}}_-$ in each region.

Similarly, for the negative sheet the magnitude of the field is

$$E_- = \frac{|\sigma_-|}{2\epsilon_0} = \frac{4.3 \times 10^{-6}\ \mathrm{C/m^2}}{(2)(8.85 \times 10^{-12}\ \mathrm{C^2/N \cdot m^2})} = 2.43 \times 10^5\ \mathrm{N/C}.$$

Figure 27-13*a* shows these fields to the left of the sheets, between them, and to the right of the sheets.

The resultant fields in these three regions follow from the vector sums of $\vec{\mathbf{E}}_+$ and $\vec{\mathbf{E}}_-$. To the left of the sheets, we have (taking components of $\vec{\mathbf{E}}$ in Fig. 27-13 to be positive if $\vec{\mathbf{E}}$ points to the right and negative if $\vec{\mathbf{E}}$ points to the left)

$$\begin{aligned} E_L = -E_+ + E_- &= -3.84 \times 10^5\ \mathrm{N/C} + 2.43 \times 10^5\ \mathrm{N/C} \\ &= -1.4 \times 10^5\ \mathrm{N/C}. \end{aligned}$$

The resultant (negative) electric field in this region points to the left, as Fig. 27-13*b* shows. To the right of the sheets, the electric field has this same magnitude but points to the right in Fig. 27-13*b*.

Between the sheets, the two fields add to give

$$\begin{aligned} E_C = E_+ + E_- &= 3.84 \times 10^5\ \mathrm{N/C} + 2.43 \times 10^5\ \mathrm{N/C} \\ &= 6.3 \times 10^5\ \mathrm{N/C}. \end{aligned}$$

Outside the sheets, the electric field behaves like that due to a single sheet whose surface charge density is $\sigma_+ + \sigma_-$ or $+2.5 \times 10^{-6}\ \mathrm{C/m^2}$. The field pattern of Fig. 27-13*b* bears this out. In Exercises 14 and 15 you can investigate the case in which the two surface charge densities are equal in magnitude but opposite in sign and also the case in which they are equal in both magnitude and sign.

27-6 GAUSS' LAW AND CONDUCTORS

We have seen that by using Gauss' law we can find the electric field for several highly symmetric charge distributions. We can also use Gauss' law to deduce the properties of conductors carrying a net electric charge. One such property is

> *An excess charge placed on an isolated conductor moves entirely to the outer surface of the conductor. None of the excess charge is found within the body of the conductor.*

Let us review what occurs when we place a quantity of electric charge on an isolated conductor. These charges can in principle be deposited anywhere in the conductor, even deep within its interior. Initially there is an electric field in the interior of the conductor due to the charges. This electric field results in forces on the charges that cause them to redistribute themselves. Very quickly (within 10^{-9} s) the electric field becomes zero, and the charges stop moving. This is the condition we describe as *electrostatic equilibrium.* If the field in the interior were nonzero, the conduction electrons in the metal would experience a force, and moving charges (an electric current) would be observed. Since we do not observe such currents, we conclude that the electric field is zero in the interior.

Keep in mind that here we are considering only an "isolated" conductor—that is, a conductor that is free from all external influences. A wire carrying a current cannot be considered an isolated conductor, because it must be connected to an external agent such as a battery. The electric field in such a wire is not zero, the wire is *not* in electrostatic equilibrium, and the conclusions of this section do not apply to the wire.

If we accept that the electric field in the interior of the conductor is zero under electrostatic conditions, then Gauss' law directly implies that the charge on the conductor must reside on its outside surface. Figure 27-14*a* shows a conductor of arbitrary shape, perhaps a lump of copper, carrying a net charge q and hanging from an insulating thread. A Gaussian surface has been drawn just inside the outer surface of the conductor.

If the electric field is zero everywhere inside the conductor, it is zero everywhere on our Gaussian surface, which lies entirely inside the conductor. This means that the flux through the Gaussian surfaces is zero. Gauss' law then allows us to conclude that the net charge enclosed by the Gaussian surface must be zero. If there is no charge inside the Gaussian surface, it must be outside that surface, which means that the charge must be on the actual outer surface of the conductor.

Why is the electric field zero inside the conductor? Suppose that we could somehow "freeze" the charges on the surface, perhaps by embedding them in a thin plastic coating, while we removed the conductor completely, leaving only a thin shell of charge. The electric field would not change at all—it would remain zero everywhere inside the shell. This shows that the electric field is set up by the charges and not by the conductor. The conductor merely provides a pathway so that the charges can easily move to take positions where they set up a net electric field of zero inside the conductor.

The Charge on Interior Surfaces

So far we have been considering the charge on the outer surface of a solid conductor. Suppose the conductor has an internal cavity, as shown in Fig. 27-14*b*. Will charge also

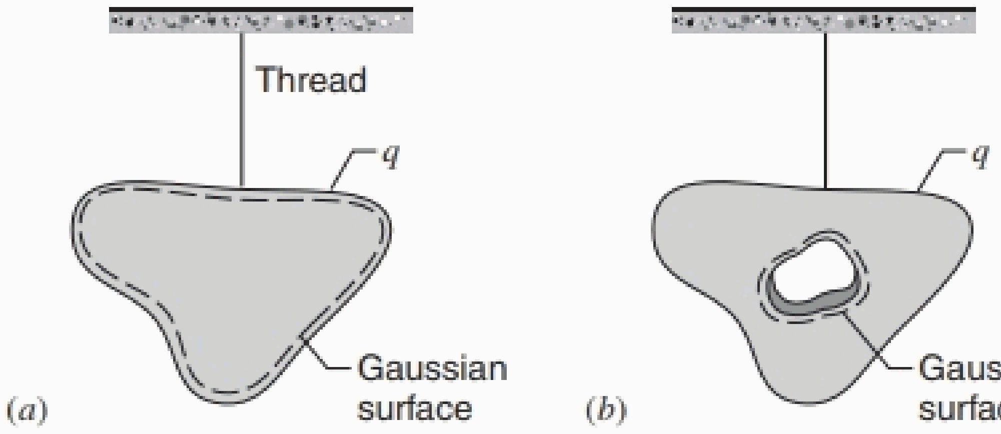


FIGURE 27-14. (*a*) An isolated metallic conductor carrying a charge q hangs from a thread. A Gaussian surface has been drawn just inside the surface of the conductor. (*b*) An internal cavity in the conductor is surrounded by a different Gaussian surface.

appear on the surface of this cavity? It is reasonable to suppose that scooping out electrically neutral material to form the cavity should not change the distribution of charge on the outer surface or the electric field in the interior. We can use Gauss' law for a quantitative proof.

We draw a Gaussian surface surrounding the cavity, close to its surface but inside the conductor, as shown in Fig. 27-14*b*. Because $\vec{\mathbf{E}} = 0$ everywhere inside the conductor, there can be no flux through this Gaussian surface. Therefore, according to Gauss' law, the surface can enclose no net charge, and so there can be no charge on the surface of the interior cavity for an isolated conductor.

If an object with a charge q' is placed inside the cavity (so that we can no longer regard our conductor as isolated), Gauss' law still requires that the net charge inside the Gaussian surface is zero. In this case, a charge $-q'$ must be attracted to the surface of the cavity to keep the net charge zero within the Gaussian surface. If the outer conductor originally carried a net charge q, then a charge of $q + q'$ will appear on its outer surface, so that the net charge does not change.

The Electric Field Outside the Conductor

Although the excess charge on an isolated conductor moves entirely to its surface, that charge—except for an isolated spherical conductor—does not in general distribute itself uniformly over that surface. Put another way, the surface charge density σ $(= dq/dA)$ varies from point to point over the surface.

We can use Gauss' law to find a relation—at any surface point—between the surface charge density σ at that point and the electric field $\vec{\mathbf{E}}$ just outside the surface at that same point. Figure 27-15*a* shows a squat cylindrical Gaussian surface, the (small) area of its two end caps being A. The end caps are parallel to the surface, one lying entirely inside the conductor and the other entirely outside. The short cylindrical walls are perpendicular to the surface of the conductor. An enlarged view of the Gaussian surface is shown in Fig. 27-15*b*.

The electric field just outside a charged isolated conductor in electrostatic equilibrium must be at right angles to the surface of the conductor. If this were not so, there would be a component of $\vec{\mathbf{E}}$ lying in the surface and this component would set up surface currents that would redistribute the surface charges, violating our assumption of electrostatic equilibrium. Thus $\vec{\mathbf{E}}$ is perpendicular to the surface of the conductor, and the flux through the exterior end cap of the Gaussian surface of Fig. 27-15*b* is EA. The flux through the interior end cap is zero, because $\vec{\mathbf{E}} = 0$ for all interior points of the conductor. The flux through the cylindrical walls is also zero because the lines of $\vec{\mathbf{E}}$ are parallel to the surface, so they cannot pierce it. The charge q enclosed by the Gaussian surface is σA.

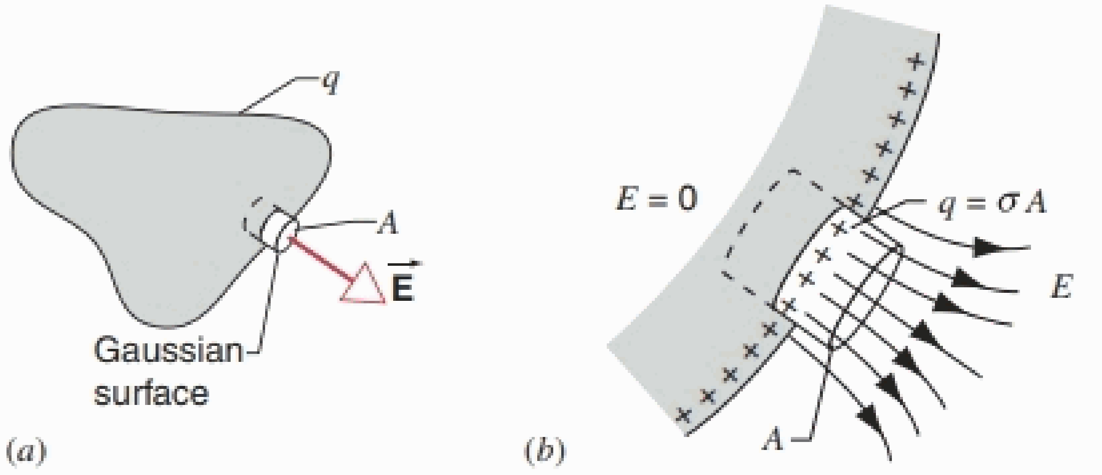


FIGURE 27-15. (*a*) A small Gaussian surface has been placed at the surface of a charged conductor. (*b*) An enlarged view of the Gaussian surface, which encloses a charge q equal to σA.

The total flux can then be calculated as

$$\begin{aligned}\Phi_E &= \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} \\ &= \int_{\text{outer cap}} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} + \int_{\text{inner cap}} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} + \int_{\text{side walls}} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} \\ &= EA + 0 + 0 = EA.\end{aligned}$$

The electric field can now be found by using Gauss' law:

$$\epsilon_0 \Phi_E = q,$$

and substituting the values for the flux and the enclosed charge q $(= \sigma A)$, we find

$$\epsilon_0 EA = \sigma A$$

or

$$E = \frac{\sigma}{\epsilon_0}. \tag{27-18}$$

Compare this result with Eq. 27-12 for the electric field near a sheet of charge: $E = \sigma/2\epsilon_0$. The electric field near a conductor is *twice* the field we would expect if we considered the conductor to be a sheet of charge, even for points very close to the surface, where the immediate vicinity *does* look like a sheet of charge. How can we understand the difference between the two cases?

A sheet of charge can be constructed by spraying charges on one side of a thin layer of plastic. The charges stick where they land and are not free to move. We cannot charge a conductor in the same way. We can imagine the surface of the conductor to be divided into two sections: the region near where we wish to find the electric field and the remainder of the conductor. If we are sufficiently close to the conductor in Fig. 27-15, the region near the Gaussian surface can be approximated as a sheet of charge, and it contributes an amount $E = \sigma/2\epsilon_0$ to the electric field. However, the charge on the rest of the conductor can be shown to contribute an identical amount to the electric field. The total electric field is the sum of the two contributions, or σ/ϵ_0.

We can see this most directly in the case of a thin conducting plate. Suppose the plate has area A. If we spray charge q anywhere on the plate, the charge will distribute itself over *both* surfaces of the plate, as in Fig. 27-16. We

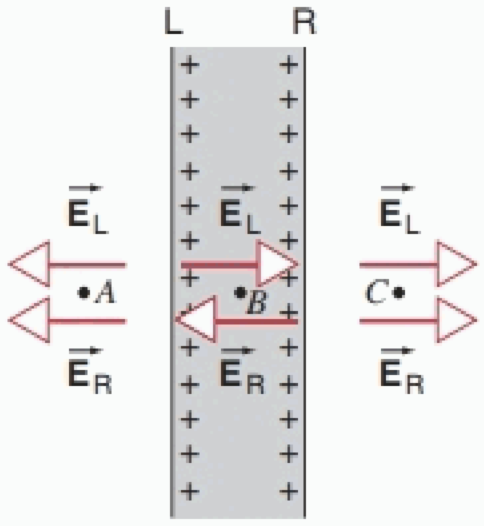


FIGURE 27-16. The electric charge near a thin conducting sheet. Note that both surfaces have charges on them. The fields $\vec{\mathbf{E}}_L$ and $\vec{\mathbf{E}}_R$ due, respectively, to the charges on the left and right surfaces reinforce at points *A* and *C*, and they cancel at points *B* in the interior of the sheet.

therefore expect to find charge $q/2$ and a charge density $\sigma = q/2A$ on each surface of the plate. We can consider each surface of the plate as a sheet of charge, which (according to Eq. 27-12) establishes an electric field $E = \sigma/2\epsilon_0 = q/4A\epsilon_0$. Close to the plate (at points *A* or *C* in Fig. 27-16) the fields due to the left and right surfaces are equal, and they add to give a total electric field of $E = q/4A\epsilon_0 + q/4A\epsilon_0 = q/2A\epsilon_0 = \sigma/\epsilon_0$. In the interior of the plate (point *B*), the fields are in the opposite direction and sum to zero, as expected for the interior of a conductor.

Suppose that now we bring a second plate, carrying a charge $-q$, to the vicinity of the first plate. Now the original conductor can no longer be considered as "isolated," and the charge on its outer surface is no longer uniformly distributed. There is an attraction between the positive charges on one plate and the negative charges on the other that draws the charges to the surfaces of the plates facing one another (Fig. 27-17). Each surface has charge q (instead of $q/2$) and charge density $\sigma = q/A$. Regarded as a sheet of charge, each surface sets up an electric field $E = \sigma/2\epsilon_0 = q/2A\epsilon_0$, according to Eq. 27-12. In the region between them, the positive and negative plates give contributions to the electric field of equal magnitudes and identical directions, so the net electric field between the plates is $E = \sigma/\epsilon_0 = q/A\epsilon_0$. This is the electric field of a parallel-plate capacitor, which is what is shown in Fig. 27-17.

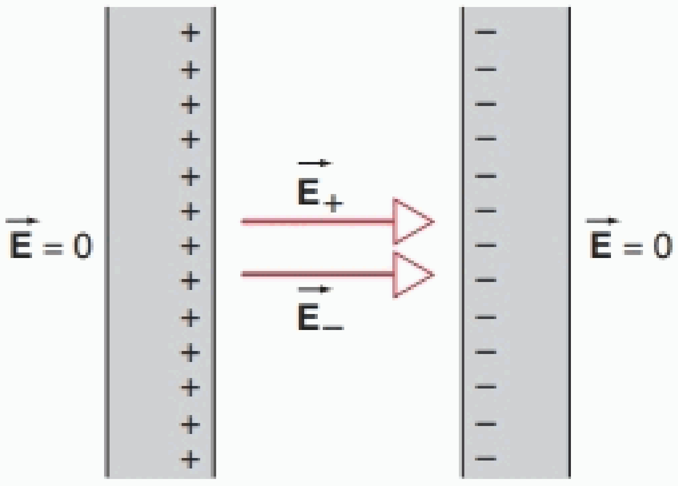


FIGURE 27-17. Two thin conducting plates carry equal and opposite charges. $\vec{\mathbf{E}}_+$ is the field due to the positively charged plate, and $\vec{\mathbf{E}}_-$ is the field set up by the negatively charged plate.

SAMPLE PROBLEM 27-4. The electric field just above the surface of the charged drum of a photocopying machine has a magnitude E of 2.3×10^5 N/C. What is the surface charge density on the drum if it is a conductor?

Solution From Eq. 27-18 we have

$$\begin{aligned}\sigma = \epsilon_0 E &= (8.85 \times 10^{-12}\ \mathrm{C^2/N\cdot m^2})(2.3 \times 10^5\ \mathrm{N/C})\\ &= 2.0 \times 10^{-6}\ \mathrm{C/m^2} = 2.0\ \mu\mathrm{C/m^2}.\end{aligned}$$

SAMPLE PROBLEM 27-5. The magnitude of the average electric field normally present in the Earth's atmosphere just above the surface of the Earth is about 150 N/C, directed downward (radially inward, toward the center of the Earth). What is the total net surface charge carried by the Earth? Assume the Earth to be a conductor.

Solution Lines of force terminate on negative charges so that, if the Earth's electric field points downward, its average surface charge density σ must be negative. From Eq. 27-18 we find

$$\begin{aligned}\sigma = \epsilon_0 E &= (8.85 \times 10^{-12}\ \mathrm{C^2/N\cdot m^2})(-150\ \mathrm{N/C})\\ &= -1.33 \times 10^{-9}\ \mathrm{C/m^2}.\end{aligned}$$

The Earth's total charge q is the surface charge density multiplied by $4\pi R^2$, the surface area of the (presumed spherical) Earth. Thus

$$\begin{aligned}q &= \sigma 4\pi R^2\\ &= (-1.33 \times 10^{-9}\ \mathrm{C/m^2})(4\pi)(6.37 \times 10^6\ \mathrm{m})^2\\ &= -6.8 \times 10^5\ \mathrm{C} = -680\ \mathrm{kC}.\end{aligned}$$

SAMPLE PROBLEM 27-6. A long hollow cylindrical conductor (inner radius a, outer radius b) is surrounded by a long coaxial cylindrical conducting shell (inner radius c, outer radius d), as shown in Fig. 27-18. The inner conductor carries a positive charge $2q$, and the outer conductor carries a charge $-3q$. Find the charge that resides on each surface of the two conductors.

Solution Gauss' law gives similar results in the cylindrical and spherical geometries. In particular, the electric field due to the outer conductor in the region $r < c$ is zero, just as we proved for the spherical shell of charge. The charges on the outer conductor produce no electric field at the location of the inner conductor, which can thus be regarded as "isolated" for this discussion. If we

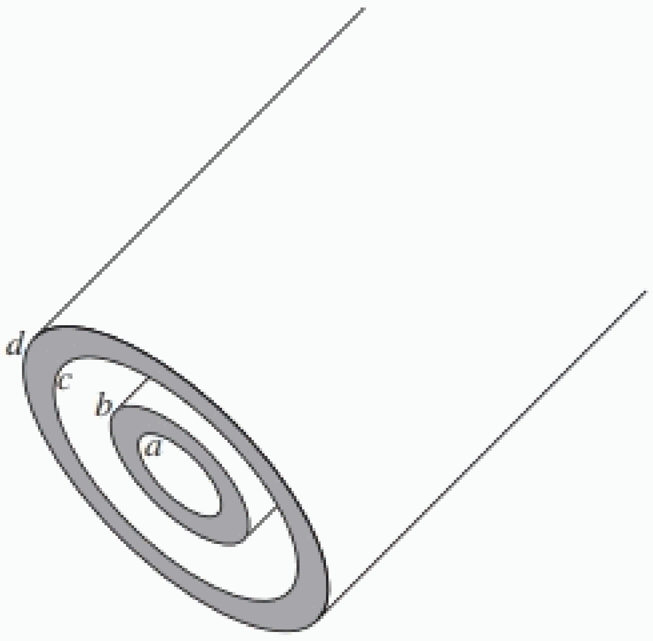


FIGURE 27-18. Sample Problem 27-6. Two coaxial cylindrical conducting shells.

treat the inner conductor as isolated, we conclude that charge must reside entirely on its outer surface. Thus there is no charge on surface a and a positive charge $2q$ on surface b.

If we were to draw a coaxial cylindrical Gaussian surface through the interior of the outer cylinder ($c < r < d$), we can use Gauss' law to conclude that the flux through the Gaussian surface is zero. The flux through the curved part of the Gaussian surface is zero, because $E = 0$ everywhere inside the conductor, and the flux through the flat ends of the surface is also zero, because the field for $b < r < c$ must be radial and therefore parallel to the flat surfaces. This means, according to Gauss' law, that the total charge inside the Gaussian surface must be zero. We know that there is a charge $2q$ on the inner conductor, so to make the net charge zero there must be a charge $-2q$ on surface c. Since the total charge on the outer cylinder is $-3q$, the remaining charge of $-q$ must appear on surface d.

Note that the outer conductor is influenced by the charge on the inner conductor and cannot be considered as isolated, so the charge on that conductor does not all reside on its outer surface.

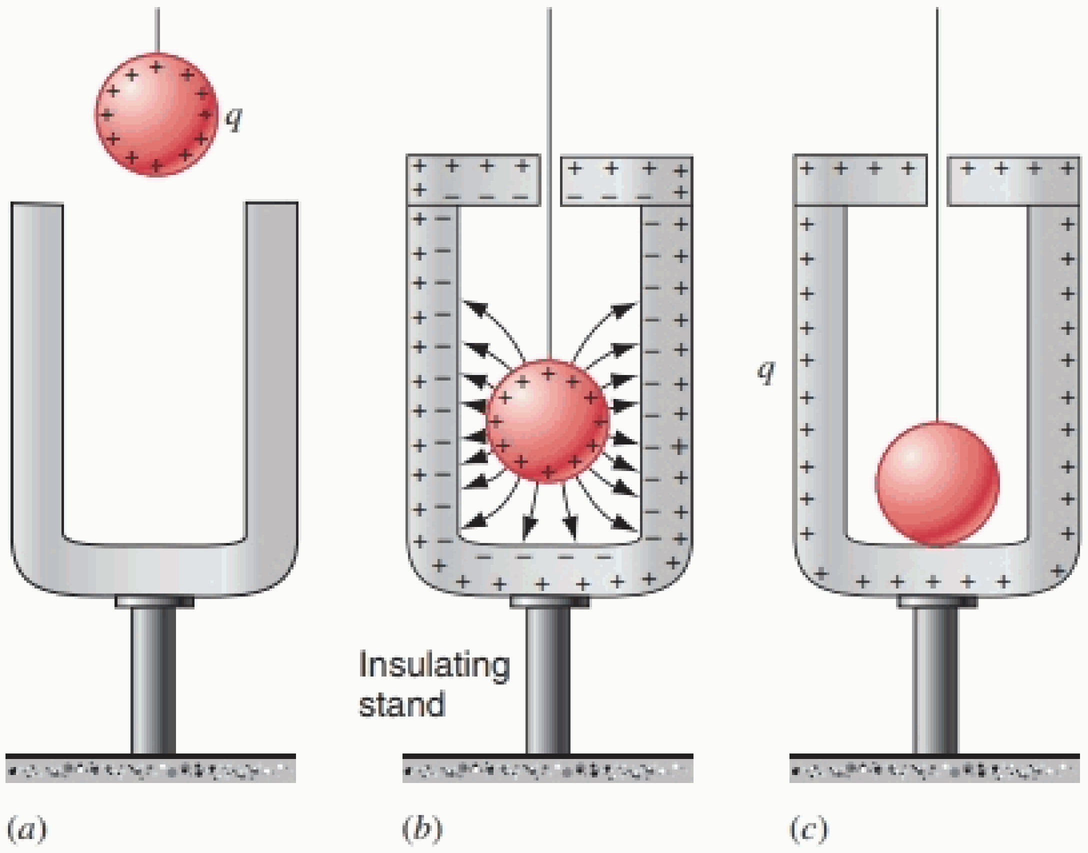


FIGURE 27-19. An arrangement conceived by Benjamin Franklin to show that the charge placed on a conductor moves to its surface. (a) A charged metal ball is lowered into an uncharged metal can. (b) The ball is inside the can and a cover is added. The field lines between the ball and the uncharged can are shown. The ball attracts charges of the opposite sign to the inside of the can. (c) When the ball touches the can, they form a single conductor, and the net charge flows to the outer surface. The ball can then be removed from the can and shown to be completely uncharged, thus proving that the charge must have been transferred entirely to the can.

27-7 EXPERIMENTAL TESTS OF GAUSS' LAW AND COULOMB'S LAW

In Section 27-6, we deduced that the excess charge in a conductor must lie only on its outer surface. No charge can be within the volume of the conductor or on the surface of an empty inner cavity. This result was derived directly from Gauss' law. Therefore, testing whether the charge does in fact lie entirely on the outer surface is a way of testing Gauss' law. If charge is found to be within the conductor or on an interior surface (such as the cavity in Fig. 27-14b), then Gauss' law fails. We also proved in Section 27-4 that Coulomb's law follows directly from Gauss' law. Thus if Gauss' law fails, then Coulomb's law fails. In particular, the force law might not be exactly an inverse square law. The exponent of r might differ from 2 by some small amount δ, so that the radial electric field might be

$$E_r = \frac{1}{4\pi\epsilon_0}\frac{q}{r^{2+\delta}}, \tag{27-19}$$

in which δ is exactly zero if Coulomb's law and Gauss' law hold.

The direct measurement of the force between two charges, described in Chapter 25, does not have the precision necessary to test whether δ is zero beyond a few percent. The observation of the charge within a conductor provides the means for a test that, as we shall see, is far more precise.

In principle, the experiment follows a procedure illustrated in Fig. 27-19. A charged metal ball hangs from an insulating thread and is lowered into a metal can that rests on an insulating stand. When the ball is touched to the inside of the can, the two objects form a *single conductor,* and, if Gauss' law is valid, all the charge from the ball must go to the outside of the combined conductor, as in Fig. 27-19c. When the ball is removed, it should no longer carry any charge. Touching other insulated metal objects to the inside of the can should not result in the transfer of any charge to the objects. Only on the outside of the can will it be possible to transfer charge.

Benjamin Franklin seems to have been the first to notice that there can be no charge inside an insulated metal can. In 1755 he wrote to a friend:

> I electrified a silver pint cann, on an electric stand, and then lowered into it a cork-ball, of about an inch in diameter, hanging by a silk string, till the cork touched the bottom of the cann. The cork was not attracted to the inside of the cann as it would have been to the outside, and though it touched the bottom, yet when drawn out, it was not found to be electrified by that touch, as it would have been by touching the outside. The fact is singular. You require the reason; I do not know it. . . .

About 10 years later Franklin recommended this "singular fact" to the attention of his friend Joseph Priestley (1733–1804). In 1767 (about 20 years before Coulomb's experiments) Priestley checked Franklin's observation and, with remarkable insight, realized that the inverse square law of force followed from it. Thus the indirect approach is not only more accurate than the direct approach of Section 25-4 but was also carried out earlier.

Priestley, reasoning by analogy with gravitation, said that the fact that no electric force acted on Franklin's cork ball when it was surrounded by a deep metal can is similar to the fact (see Section 14-5) that no gravitational force acts on a particle inside a spherical shell of matter; if gravitation obeys an inverse square law, perhaps the electrical force does also. Considering Franklin's experiment, Priestley reasoned:

> May we not infer from this that the attraction of electricity is subject to the same laws with that of gravitation and is therefore according to the squares of the distances; since it is easily demonstrated that were the earth in the form of a shell, a body in the inside of it would not be attracted to one side more than another?

Note how knowledge of one subject (gravitation) helps in understanding another (electrostatics).

Michael Faraday also carried out experiments designed to show that excess charge resides on the outside surface of a conductor. In particular, he built a large metal-covered box, which he mounted on insulating supports and charged with a powerful electrostatic generator. In Faraday's words:

> I went into the cube and lived in it, and using lighted candles, electrometers, and all other tests of electrical states, I could not find the least influence upon them . . . though all the time the outside of the cube was very powerfully charged, and large sparks and brushes were darting off from every part of its outer surface.

Coulomb's law is vitally important in physics, and if δ in Eq. 27-19 is not zero, there are serious consequences for our understanding of electromagnetism and quantum physics. The best way to measure δ is to find out *by experiment* whether an excess charge, placed on an isolated conductor, does or does not move *entirely* to its outside surface.

Modern experiments, carried out with remarkable precision, have shown that if δ in Eq. 27-19 is not zero it is certainly very, very small. Table 27-1 summarizes the results of the most important of these experiments.

Figure 27-20 is a drawing of the apparatus used by Plimpton and Lawton to measure δ. It consists in principle of two concentric metal shells, A and B, the former being 1.5 m in diameter. The inner shell contains a sensitive electrometer E connected so that it will indicate whether any charge moves between shells A and B. If the shells are connected electrically, any charge placed on the shell assembly should reside entirely on shell A if Gauss' law—and thus Coulomb's law—is correct as stated.

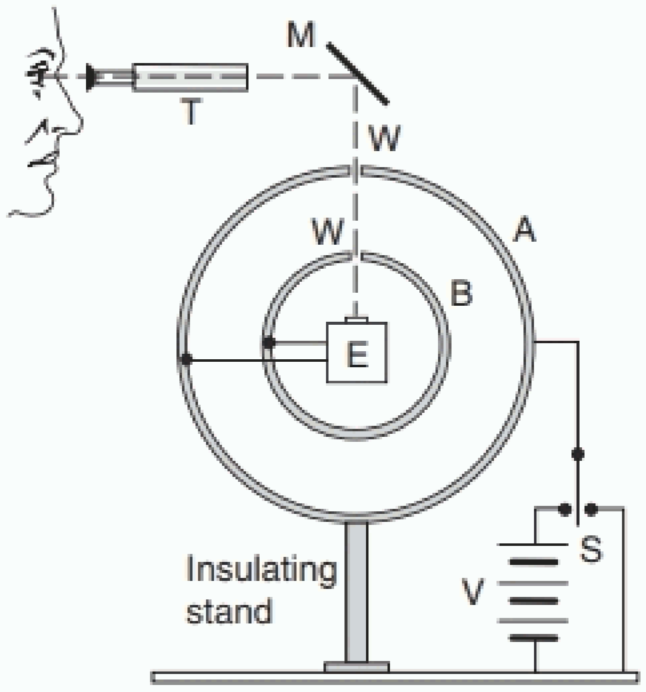


FIGURE 27-20. A modern and more precise version of the apparatus of Fig. 27-19, also designed to verify that charge resides only on the outside surface of a conductor. Charge is placed on sphere A by throwing switch S to the left, and the sensitive electrometer E is used to search for any charge that might move to the inner sphere B. It is expected that all the charge will remain on the outer surface (sphere A).

By throwing switch S to the left, a substantial charge could be placed on the sphere assembly by the battery V. If any of this charge moved to shell B, it would have to pass through the electrometer and would cause a deflection, which could be observed optically using telescope T, mirror M, and windows W.

However, when the switch S was thrown alternately from left to right, connecting the shell assembly either to the battery or to the ground, no effect was observed. Knowing the sensitivity of their electrometer, Plimpton and Lawton calculated that δ in Eq. 27-19 differs from zero by no more than 2×10^{-9}, a very small value indeed. Yet since their experiment, the limits on δ have been improved by more than seven orders of magnitude by experimenters using more detailed and precise versions of this basic apparatus.

TABLE 27-1 Tests of Coulomb's Inverse Square Law

Experimenters	*Date*	*δ (Eq. 27-19)*
Franklin	1755	
Priestley	1767	. . . according to the squares . . .
Robison	1769	<0.06
Cavendish	1773	<0.02
Coulomb	1785	a few percent at most
Maxwell	1873	$<5 \times 10^{-5}$
Plimpton and Lawton	1936	$<2 \times 10^{-9}$
Bartlett, Goldhagen, and Phillips	1970	$<1.3 \times 10^{-13}$
Williams, Faller, and Hill	1971	$<1.0 \times 10^{-16}$

MULTIPLE CHOICE

27-1 What Is Gauss' Law All About?

27-2 The Flux of a Vector Field

1. A velocity field $\vec{v}$ exists in a region of space. A closed surface S is divided into four sections, S_1, S_2, S_3, and S_4. There is a source located just outside the closed surface near S_1; there may be other sources or sinks near the other surfaces S_n, but none are inside S.

(*a*) What can be concluded about Φ_1, the flux through S_1?

(A) $\Phi_1 > 0$ (B) $\Phi_1 = 0$ (C) $\Phi_1 < 0$
(D) Nothing can be concluded about Φ_1 without additional information.

(*b*) Which of the following statements is correct about the flux through the four surfaces?

(A) At least one of the Φ_n must be negative.
(B) At least one of the Φ_n must be positive.
(C) At least one of the Φ_n must be zero.
(D) If A is correct, then so is B.
(E) Either A or B could be correct, but not both.

(*c*) Measurements indicate that $\Phi_1 + \Phi_2 > 0$. From this information we can conclude that

(A) $\Phi_3 = \Phi_4$. (B) $\Phi_3 = -\Phi_4$. (C) $\Phi_3 > \Phi_4$.
(D) $\Phi_3 < -\Phi_4$.

27-3 The Flux of the Electric Field

2. The flux through a flat surface of area A in a uniform $\vec{E}$ field is a maximum when

(A) the surface is parallel to $\vec{E}$.
(B) the surface is perpendicular to $\vec{E}$.
(C) the surface is shaped like a rectangle.
(D) the surface is shaped like a square.

3. A closed spherical surface of radius a is in a uniform electric field $\vec{E}$. What is the electric flux Φ_E through the surface?

(A) $\Phi_E = 4\pi a^2 E$ (B) $\Phi_E = \pi a^2 E$
(C) $\Phi_E = 0$
(D) Φ_E cannot be determined without additional knowledge.

27-4 Gauss' Law

4. Consider two concentric spherical surfaces, S_1 with radius a and S_2 with radius $2a$, both centered on the origin. There is a charge $+q$ at the origin, and no other charges. Compare the flux Φ_1 through S_1 with the flux Φ_2 through S_2.

(A) $\Phi_1 = 4\Phi_2$. (B) $\Phi_1 = 2\Phi_2$.
(C) $\Phi_1 = \Phi_2$. (D) $\Phi_1 = \Phi_2/2$.

5. An imaginary closed spherical surface S of radius R is centered on the origin. A positive charge is originally at the origin, and the flux through the surface is Φ_E. The positive charge is slowly moved from the origin to a point $R/2$ away from the origin. In doing so the flux through S

(A) increases to $4\Phi_E$. (B) increases to $2\Phi_E$.
(C) remains the same. (D) decreases to $\Phi_E/2$.
(E) decreases to $\Phi_E/4$.

6. Under what conditions can the electric flux Φ_E be found through a closed surface?

(A) If the magnitude of $\vec{E}$ is known everywhere on the surface.
(B) If the total charge inside the surface is specified.
(C) If the total charge outside the surface is specified.
(D) Only if the location of each point charge inside the surface is specified.

7. An imaginary, closed, spherical surface S of radius R is centered on the origin. A positive charge $+q$ is originally at the origin, and the flux through the surface is Φ_E. Three additional charges are now added along the x axis: $-3q$ at $x = -R/2$, $+5q$ at $x = R/2$, and $+4q$ at $x = 3R/2$. The flux through S is now

(A) $2\Phi_E$. (B) $3\Phi_E$. (C) $6\Phi_E$. (D) $7\Phi_E$.
(E) Φ_E cannot be determined, because the problem is no longer symmetric.

27-5 Applications of Gauss' Law

8. A dipole lies on the x axis, with the positive charge $+q$ at $x = +d/2$, and the negative charge at $-d/2$. The electric flux Φ_E through the yz plane midway between the charges

(A) is zero. (B) depends on d.
(C) depends on q. (D) depends on both q and d.

9. The surface in Multiple-choice question 8 is shifted closer to the positive charge. As the surface moves, the flux Φ_E through the surface

(A) increases. (B) decreases.
(C) remains the same.

10. The surface in Multiple-choice question 8 is instead *rotated* so that the normal to the surface is no longer parallel to the x axis. As the surface moves, the flux Φ_E through the surface

(A) increases. (B) decreases.
(C) remains the same.

11. In which of the following problems would Gauss' law be useful?

(A) Finding the electric field at various points on the surface of a uniformly charged cylinder of finite length
(B) Finding the electric flux through the end surface of a uniformly charged cylinder
(C) Finding the electric field at various points on the surface of a uniformly charged cube
(D) Finding the electric flux through one side of a uniformly charged cube

27-6 Gauss' Law and Conductors

12. A hollow conducting ball has a single positive charge $+q$ fixed at the center. The ball has no net charge.

(*a*) The charge on the inner surface of the ball is

(A) $+2q$. (B) $+q$. (C) $-q$. (D) 0.

(*b*) The charge on the outer surface of the ball is

(A) $+2q$. (B) $+q$. (C) $-q$. (D) 0.

13. Suppose that a net charge $+q$ is placed on the ball of Multiple-choice question 12; the point charge remains at its center.

(*a*) The charge on the inner surface of the ball is

(A) $+2q$. (B) $+q$. (C) $-q$. (D) 0.

(*b*) The charge on the outer surface of the ball is

(A) $+2q$. (B) $+q$. (C) $-q$. (D) 0.

14. The positive charge at the center of the ball in Multiple-choice question 13 is moved off center closer to the inner surface, but it does not touch the inner surface.

(*a*) The total charge on the inner surface of the ball will

(A) increase. (B) decrease.
(C) remain the same.
(D) change, depending on how close the ball gets to the inner surface.

(*b*) The total charge on the outer surface of the ball will

(A) increase. (B) decrease.
(C) remain the same.
(D) change, depending on how close the ball gets to the inner surface.

27-7 Experimental Tests of Gauss' Law and Coulomb's Law

QUESTIONS

1. What is the basis for the statement that lines of electric force begin and end only on electric charges?

2. Positive charges are sometimes called "sources" and negative charges "sinks" of electric field. How would you justify this terminology? Are there sources and/or sinks of gravitational field?

3. By analogy with Φ_E, how would you define the flux Φ_g of a gravitational field? What is the flux of the Earth's gravitational field through the boundaries of a room, assumed to contain no matter? Through a spherical surface closely surrounding the Earth? Through a spherical surface the size of the Moon's orbit?

4. Consider the Gaussian surface that surrounds part of the charge distribution shown in Fig. 27-21. (*a*) Which of the charges contribute to the electric field at point *P*? (*b*) Would the value obtained for the flux through the surface, calculated using only the field due to q_1 and q_2, be greater than, equal to, or less than that obtained using the total field?

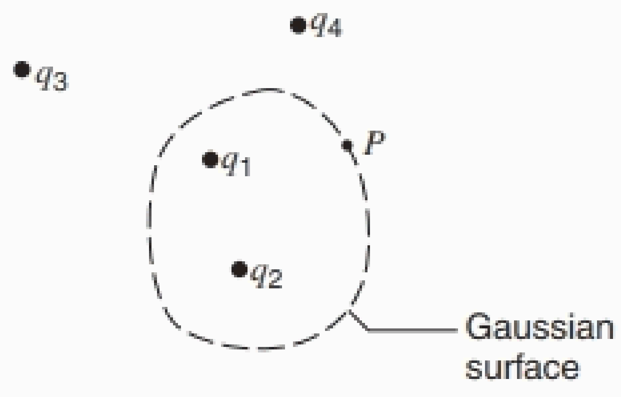


FIGURE 27-21. Question 4.

5. Suppose that an electric field in some region is found to have a constant direction but to be decreasing in strength in that direction. What do you conclude about the charge in the region? Sketch the lines of force.

6. Is it precisely true that Gauss' law states that the total number of lines of force crossing any closed surface in the outward direction is proportional to the net positive charge enclosed within the surface?

7. A point charge is placed at the center of a spherical Gaussian surface. Is Φ_E changed (*a*) if the surface is replaced by a cube of the same volume; (*b*) if the sphere is replaced by a cube of one-tenth the volume; (*c*) if the charge is moved off-center in the original sphere, still remaining inside; (*d*) if the charge is moved just outside the original sphere; (*e*) if a second charge is placed near, and outside, the original sphere; and (*f*) if a second charge is placed inside the Gaussian surface?

8. In Gauss' law $\epsilon_0 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q$, is $\vec{\mathbf{E}}$ necessarily the electric field attributable to the charge q?

9. A surface encloses an electric dipole. What can you say about Φ_E for this surface?

10. Suppose that a Gaussian surface encloses no net charge. Does Gauss' law require that $\vec{\mathbf{E}}$ equal zero for all points on the surface? Is the converse of this statement true; that is, if $\vec{\mathbf{E}}$ equals zero everywhere on the surface, does Gauss' law require that there be no net charge inside?

11. Is Gauss' law useful in calculating the field due to three equal charges located at the corners of an equilateral triangle? Explain why or why not.

12. A total charge Q is distributed uniformly throughout a cube of edge length a. Is the resulting electric field at an external point P, a distance r from the center C of the cube, given by $E = Q/4\pi\epsilon_0 r^2$? See Fig. 27-22. If not, can E be found by constructing a "concentric" cubical Gaussian surface? If not, explain why not. Can you say anything about E if $r \gg a$?

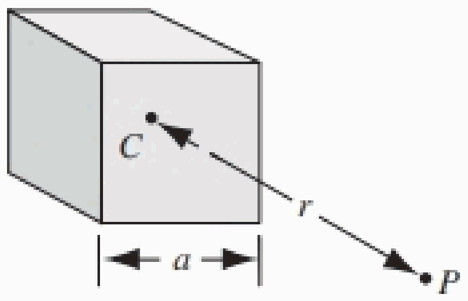


FIGURE 27-22. Question 12.

13. Is $\vec{\mathbf{E}}$ necessarily zero inside a charged rubber balloon if the balloon is (*a*) spherical or (*b*) sausage-shaped? For each shape, assume the charge to be distributed uniformly over the surface. How would the situation change, if at all, if the balloon has a thin layer of conducting paint on its outside surface?

14. A spherical rubber balloon carries a charge that is uniformly distributed over its surface. As the balloon is blown up, how does E vary for points (*a*) inside the balloon, (*b*) at the surface of the balloon, and (*c*) outside the balloon?

15. In Section 27-4 we have seen that Coulomb's law can be derived from Gauss' law. Does this necessarily mean that Gauss' law can be derived from Coulomb's law?

16. Would Gauss' law hold if the exponent in Coulomb's law were not exactly 2?

17. A large, insulated, hollow conductor carries a positive charge. A small metal ball carrying a negative charge of the same magnitude is lowered by a thread through a small opening in the top of the conductor, allowed to touch the inner surface, and then withdrawn. What then is the charge on (*a*) the conductor and (*b*) the ball?

18. Can we deduce from the argument of Section 27-6 that the electrons in the wires of a house wiring system move along the surfaces of those wires? If not, why not?

19. In Section 27-6, we assumed that $\vec{\mathbf{E}}$ equals zero everywhere inside an isolated conductor. However, there are certainly very large electric fields inside the conductor, at points close to the electrons or to the nuclei. Does this invalidate the proof of Section 27-4? Explain.

20. Does Gauss' law, as applied in Section 27-6, require that all the conduction electrons in an insulated conductor reside on the surface?

21. A positive point charge q is located at the center of a hollow metal sphere. What charges appear on (*a*) the inner surface and (*b*) the outer surface of the sphere? (*c*) If you bring an (uncharged) metal object near the sphere, will it change your answers to (*a*) or (*b*)? Will it change the way charge is distributed over the sphere?

22. If a charge $-q$ is distributed uniformly over the surface of a thin, insulated, spherical metal shell of radius a, there will be no electric field inside. If now a point charge $+q$ is placed at the center of the sphere, there will be no external field. This point charge can be displaced a distance $d < a$ from the center, but that gives the system a dipole moment and creates an external field. How do you account for the energy appearing in this external field?

23. How can you remove completely the excess charge from a small conducting body?

24. Explain why the spherical symmetry of Fig. 27-7 restricts us to a consideration of $\vec{\mathbf{E}}$ that has only a radial component at any point. (Hint: Imagine other components, perhaps along the equivalent of longitude or latitude lines on the Earth's surface. Spherical symmetry requires that these look the same from any perspective. Can you invent such field lines that satisfy this criterion?)

25. Explain why the symmetry of Fig. 27-8 restricts us to a consideration of $\vec{\mathbf{E}}$ that has only a radial component at any point. Remember, in this case, that the field must not only look the same at any point along the line but must also look the same if the figure is turned end for end.

26. The total charge on a charged infinite rod is infinite. Why is not E also infinite? After all, according to Coulomb's law, if q is infinite, so is E.

27. Explain why the symmetry of Fig. 27-9 restricts us to a consideration of $\vec{\mathbf{E}}$ that has only a component directed away from the sheet. Why, for example, could $\vec{\mathbf{E}}$ not have a component parallel to the sheet? Remember, in this case, that the field must not only look the same at any point along the sheet in any direction but must also look the same if the sheet is rotated about a line perpendicular to the sheet.

28. The field due to an infinite sheet of charge is uniform, having the same strength at all points no matter how far from the surface charge. Explain how this can be, given the inverse square nature of Coulomb's law.

29. As you penetrate a uniform sphere of charge, E should decrease because less charge lies inside a sphere drawn through the observation point. On the other hand, E should increase because you are closer to the center of this charge. Which effect dominates, and why?

30. Given a spherically symmetric charge distribution (not of uniform radial density of charge), is E necessarily a maximum at the surface? Comment on various possibilities.

31. Does Eq. 27-15 hold true for Fig. 27-11*a* if (*a*) there is a concentric spherical cavity in the body, (*b*) a point charge Q is at the center of this cavity, and (*c*) the charge Q is inside the cavity but not at its center?

EXERCISES

27-1 What Is Gauss' Law All About?

27-2 The Flux of a Vector Field

27-3 The Flux of the Electric Field

1. The square surface shown in Fig. 27-23 measures 3.2 mm on each side. It is immersed in a uniform electric field with $E = 1800$ N/C. The field lines make an angle of 65° with the "outward pointing" normal, as shown. Calculate the flux through the surface.

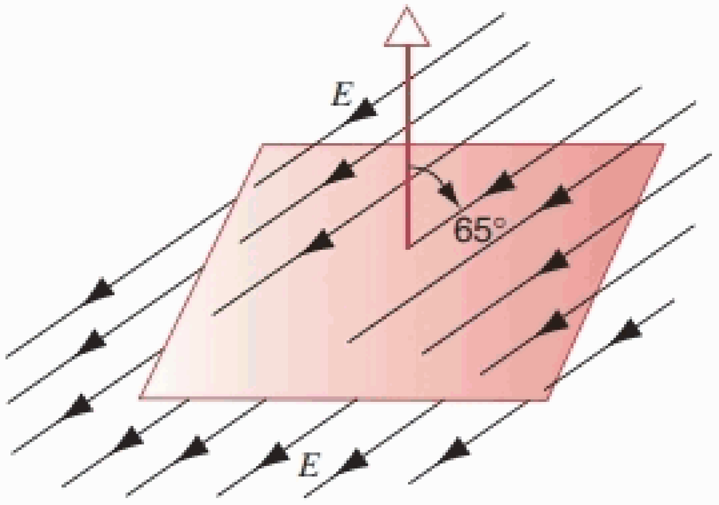


FIGURE 27-23. Exercise 1.

2. A cube with 1.4-m edges is oriented as shown in Fig. 27-24 in a region of uniform electric field. Find the electric flux through the right face if the electric field is given by (*a*) (6 N/C)$\hat{\mathbf{i}}$, (*b*) (− 2 N/C)$\hat{\mathbf{j}}$, and (*c*) (− 3 N/C)$\hat{\mathbf{i}}$ + (4 N/C)$\hat{\mathbf{k}}$. (*d*) Calculate the total flux through the cube for each of these fields.

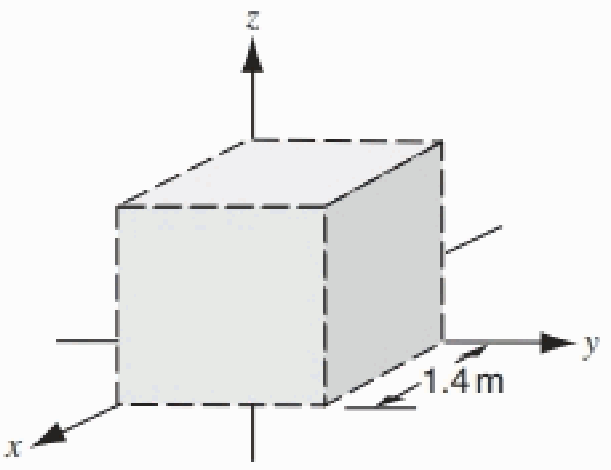


FIGURE 27-24. Exercise 2.

3. Calculate Φ_E through (*a*) the flat base and (*b*) the curved surface of a hemisphere of radius R. The field $\vec{\mathbf{E}}$ is uniform and parallel to the axis of the hemisphere, and the lines of $\vec{\mathbf{E}}$ enter through the flat base. (Use the outward pointing normal.)

27-4 Gauss' Law

4. Charge on an originally uncharged insulated conductor is separated by holding a positively charged rod very closely nearby, as in Fig. 27-25. Calculate the flux for the five Gaussian surfaces shown. Assume that the induced negative charge on the conductor is equal to the positive charge q on the rod.

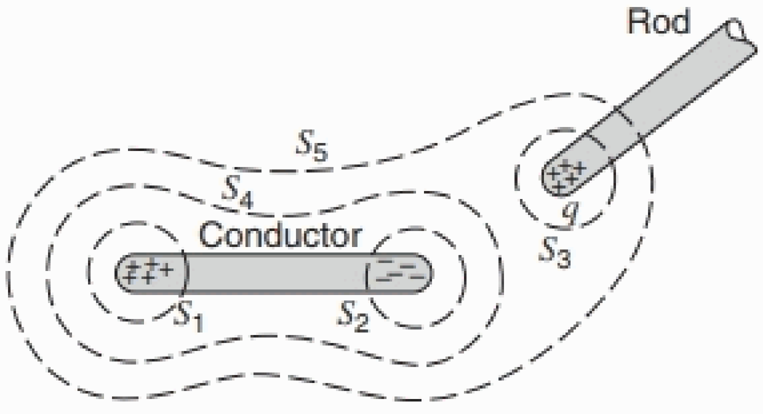


FIGURE 27-25. Exercise 4.

5. A point charge of 1.84 μC is at the center of a cubical Gaussian surface 55 cm on edge. Find Φ_E through the surface.

6. The net electric flux through each face of a die (one member of a pair of dice) has magnitude in units of 10^3 N·m^2/C equal to the number N of spots on the face (1 through 6). The flux is inward for N odd and outward for N even. What is the net charge inside the die?

7. A point charge $+q$ is a distance $d/2$ from a square surface of side d and is directly above the center of the square as shown in Fig. 27-26. Find the electric flux through the square. (Hint: Think of the square as one face of a cube with edge d.)

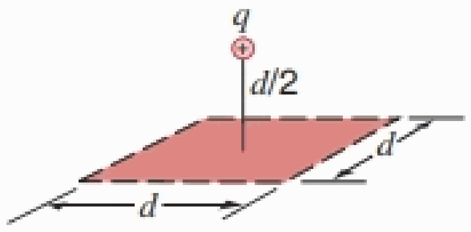


FIGURE 27-26. Exercise 7.

8. A butterfly net is in a uniform electric field E as shown in Figure 27-27. The rim, a circle of radius a, is aligned perpendicular to the field. Find the electric flux through the netting, relative to the outward normal.

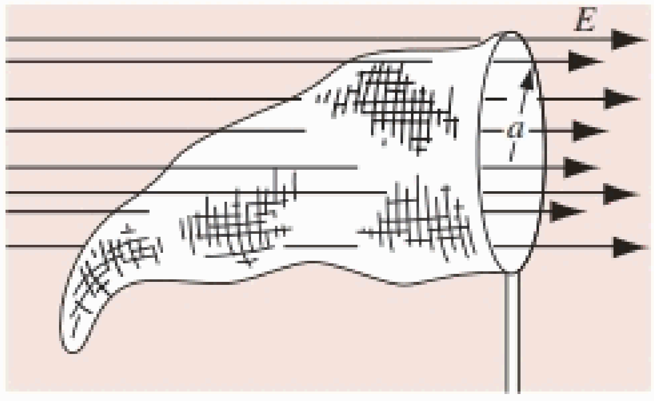


FIGURE 27-27. Exercise 8.

9. It is found experimentally that the electric field in a certain region of the Earth's atmosphere is directed vertically down. At an altitude of 300 m the field is 58 N/C and at an altitude of 200 m it is 110 N/C. Find the net amount of charge contained in a cube 100 m on edge that is located at an altitude between 200 and 300 m. Neglect the curvature of the Earth.

10. Find the net flux through the cube of Exercise 2 and Figure 27-14 if the electric field is given by (*a*) $\vec{\mathbf{E}}$ = (3 N/C·m)$y\hat{\mathbf{j}}$ and (*b*) (−4 N/C)$\hat{\mathbf{i}}$ + [6 N/C + (3 N/C·m)y]$\hat{\mathbf{j}}$. (*c*) In each case, how much charge is inside the cube?

11. A point charge q is placed at one corner of a cube of edge a. What is the flux through each of the cube faces? (Hint: Use Gauss' law and symmetry arguments.)

27-5 Applications of Gauss' Law

12. An infinite line of charge produces a field of 4.52×10^4 N/C at a distance of 1.96 m. Calculate the linear charge density.

13. (*a*) The drum of the photocopying machine in Sample Problem 27-4 has a length of 42 cm and a diameter of 12 cm. What is the total charge on the drum? (*b*) The manufacturer wishes to produce a desktop version of the machine. This requires reducing the size of the drum to a length of 28 cm and a diameter of 8.0 cm. The electric field at the drum surface must remain unchanged. What must be the charge on this new drum?

14. Two thin, large, nonconducting sheets of positive charge face each other as in Figure 27-28. What is $\vec{\mathbf{E}}$ at points (*a*) to the left of the sheets, (*b*) between them, and (*c*) to the right of the sheets? Assume the same surface charge density σ for each sheet. Consider only points not near the edges whose distance from the sheets is small compared to the dimensions of the sheets. (Hint: See Sample Problem 27-3.)

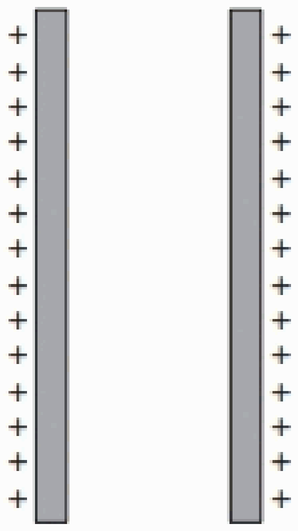


FIGURE 27-28. Exercise 14.

15. Two large metal plates face each other as in Figure 27-29 and carry charges with surface charge density $+\sigma$ and $-\sigma$, respectively, on their inner surfaces. Find E at points (*a*) to the left of the plates, (*b*) between them, and (*c*) to the right of the plates. Consider only points not near the edges whose distances from the plates are small compared to the dimensions of the plates. (Hint: See Sample Problem 27-3.)

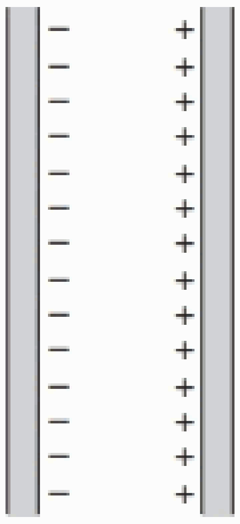


FIGURE 27-29. Exercise 15.

16. An electron remains stationary in an electric field directed downward in the Earth's gravitational field. If the electric field is due to charge on two large, parallel, conducting plates, oppositely charged and separated by 2.3 cm, what is the surface charge density, assumed to be uniform, on the plates?

17. A very long, straight, thin wire carries -3.60 nC/m of fixed negative charge. The wire is to be surrounded by a uniform cylinder of positive charge, radius 1.50 cm, coaxial with the wire. The volume charge density ρ of the cylinder is to be selected so that the net electric field outside the cylinder is zero. Calculate the required positive charge density ρ.

18. Figure 27-30 shows a point charge $q = 126$ nC at the center of a spherical cavity of radius 3.66 cm in a piece of metal. Use Gauss' law to find the electric field (*a*) at point P_1 halfway from the center to the surface, and (*b*) at point P_2.

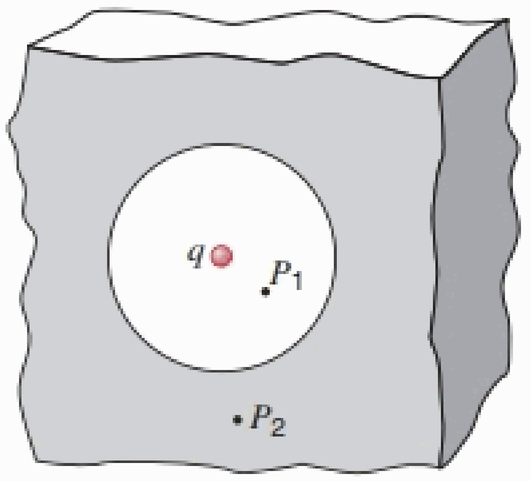


FIGURE 27-30. Exercise 18.

19. A proton orbits with a speed $v = 294$ km/s just outside a charged sphere of radius $r = 1.13$ cm. Find the charge on the sphere.

20. Two charged, concentric, thin, spherical shells have radii of 10.0 cm and 15.0 cm. The charge on the inner shell is 40.6 nC and that on the outer shell is 19.3 nC. Find the electric field (*a*) at $r = 12.0$ cm, (*b*) at $r = 22.0$ cm, and (*c*) at $r = 8.18$ cm from the center of the shells.

21. Two long, charged, concentric cylinders have radii of 3.22 and 6.18 cm. The surface charge density on the inner cylinder is 24.1 μC/m^2 and that on the outer cylinder is -18.0 μC/m^2. Find the electric field at (*a*) $r = 4.10$ cm and (*b*) $r = 8.20$ cm.

22. Positive charge is distributed uniformly throughout a long, nonconducting, cylindrical shell of inner radius R and outer radius $2R$. At what radial depth beneath the outer surface of the charge distribution is the electric field strength equal to one-half the surface value?

23. A 115-keV electron is fired directly toward a large, flat, plastic sheet that has a surface charge density of -2.08 μC/m^2. From what distance must the electron be fired if it is just to fail to strike the sheet? (Ignore relativistic effects.)

24. Construct a spherical Gaussian surface centered on an infinite line of charge, calculate the flux through the sphere, and thereby show that Gauss' law is satisfied.

25. Charge is distributed uniformly throughout an infinitely long cylinder of radius R. (*a*) Show that E at a distance r from the cylinder axis ($r < R$) is given by

$$E = \frac{\rho r}{2\epsilon_0},$$

where ρ is the volume charge density. (*b*) What result do you obtain for $r > R$?

27-6 Gauss' Law and Conductors

26. A uniformly charged conducting sphere of 1.22 m radius has a surface charge density of 8.13 μC/m^2. (*a*) Find the charge on the sphere. (*b*) What is the total electric flux leaving the surface of the sphere? (*c*) Calculate the electric field at the surface of the sphere.

27. Space vehicles traveling through the Earth's radiation belts collide with trapped electrons. Since in space there is no ground, the resulting charge buildup can become significant and can damage electronic components, leading to control-circuit upsets and operational anomalies. A spherical metallic satellite 1.3 m in diameter accumulates 2.4 μC of charge in one orbital revolution. (*a*) Find the surface charge density. (*b*) Calculate the resulting electric field just outside the surface of the satellite.

28. Equation 27-18 ($E = \sigma/\epsilon_0$) gives the electric field at points near a charged conducting surface. Apply this equation to a conducting sphere of radius r, carrying a charge q on its surface, and show that the electric field outside the sphere is the same as the field of a point charge at the position of the center of the sphere.

29. A metal plate 8.0 cm on a side carries a total charge of 6.0 μC. (*a*) Using the infinite plate approximation, calculate the electric field 0.50 mm above the surface of the plate near the plate's center. (*b*) Estimate the field at a distance of 30 m.

27-7 Experimental Tests of Gauss' Law and Coulomb's Law

PROBLEMS

1. Gauss' law for gravitation is

$$\frac{1}{4\pi G}\Phi_g = \frac{1}{4\pi G}\oint \vec{\mathbf{g}}\cdot d\vec{\mathbf{A}} = -m,$$

where m is the enclosed mass and G is the universal gravitation constant. Derive Newton's law of gravitation from this. What is the significance of the negative sign?

2. The electric field components in Fig. 27-31 are $E_y = by^{1/2}$, $E_x = E_z = 0$, in which $b = 8830\ \text{N/C}\cdot\text{m}^{1/2}$. Calculate (*a*) the flux Φ_E through the cube and (*b*) the charge within the cube. Assume that $a = 13.0$ cm.

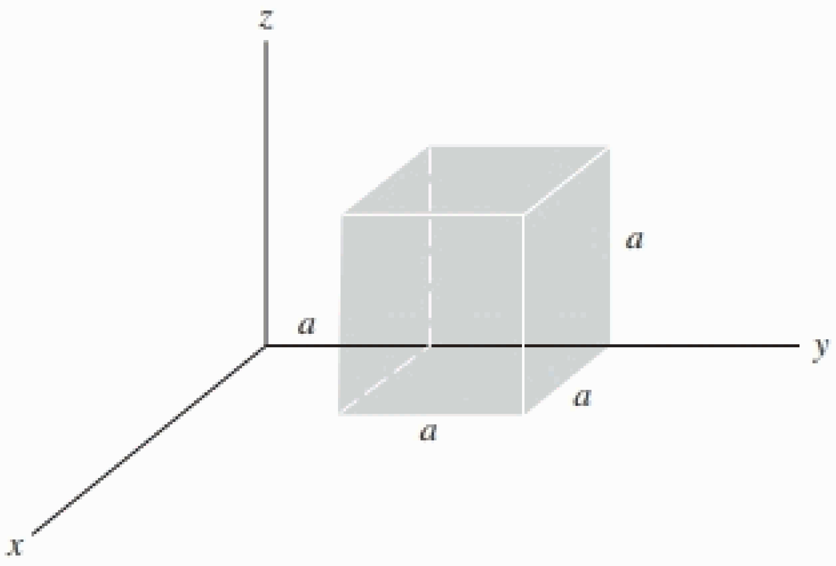


FIGURE 27-31. Problem 2.

3. A small sphere whose mass m is 1.12 mg carries a charge q = 19.7 nC. It hangs in the Earth's gravitational field from a silk thread that makes an angle $\theta = 27.4°$ with a large, uniformly charged, nonconducting sheet as in Fig. 27-32. Calculate the uniform charge density σ for the sheet.

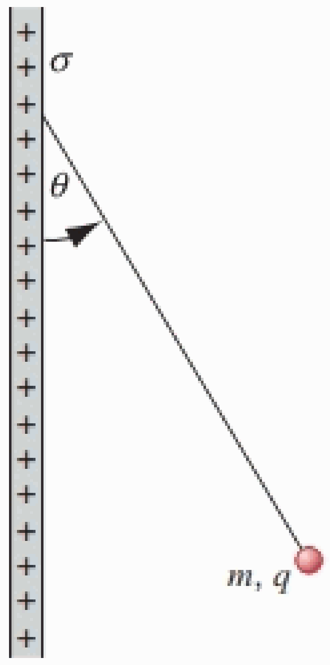


FIGURE 27-32. Problem 3.

4. Figure 27-33 shows a charge $+q$ arranged as a uniform conducting sphere of radius a and placed at the center of a spherical conducting shell of inner radius b and outer radius c. The outer shell carries a charge of $-q$. Find $E(r)$ at locations (*a*) within the sphere ($r < a$), (*b*) between the sphere and the shell ($a < r < b$), (*c*) inside the shell ($b < r < c$), and (*d*) outside the shell ($r > c$). (*e*) What charges appear on the inner and outer surfaces of the shell?

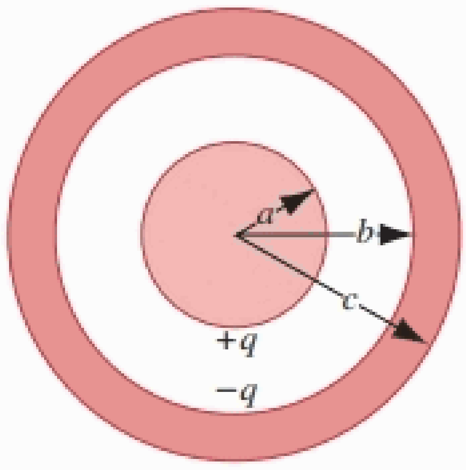


FIGURE 27-33. Problem 4.

5. A very long conducting cylinder (length L) carrying a total charge $+q$ is surrounded by a conducting cylindrical shell (also of length L) with total charge $-2q$, as shown in cross section in Fig. 27-34. Use Gauss' law to find (*a*) the electric field at points outside the conducting shell, (*b*) the distribution of the charge on the conducting shell, and (*c*) the electric field in the region between the cylinders.

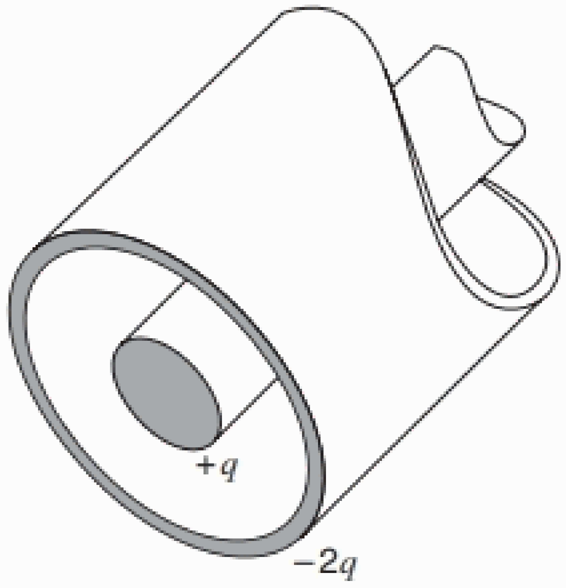


FIGURE 27-34. Problem 5.

6. A large, flat, nonconducting surface carries a uniform charge density σ. A small circular hole of radius R has been cut in the middle of the sheet, as shown in Fig. 27-35. Ignore fringing of the field lines around all edges and calculate the electric field at point P, a distance z from the center of the hole along its axis. (Hint: See Eq. 26-19 and use the principle of superposition.)

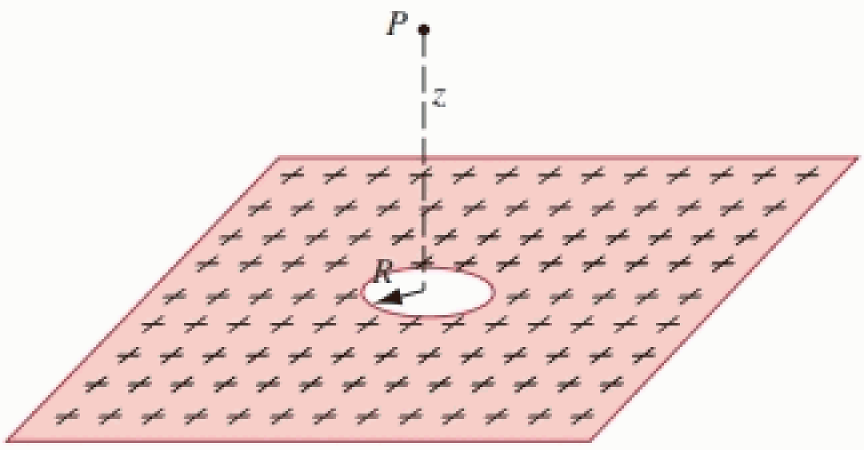


FIGURE 27-35. Problem 6.

7. Figure 27-36 shows a section through a long, thin-walled metal tube of radius R, carrying a charge per unit length λ on its surface. Derive expressions for E for various distances r from the tube axis, considering both (*a*) $r > R$ and (*b*) $r < R$. (*c*) Plot your results for the range $r = 0$ to $r = 5.0$ cm, assuming that $\lambda = 2.0 \times 10^{-8}$ C/m and $R = 3.0$ cm. (Hint: Use cylindrical Gaussian surfaces, coaxial with the metal tube.)

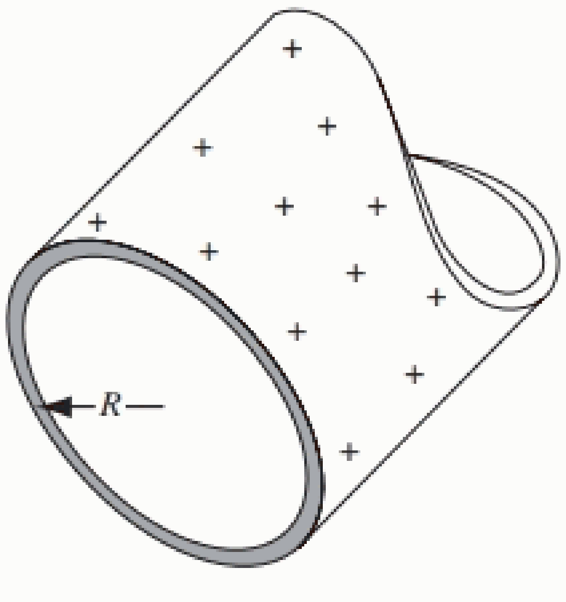


FIGURE 27-36. Problem 7.

8. Figure 27-37 shows a section through two long, thin, concentric cylinders of radii a and b. The cylinders carry equal and opposite charges per unit length λ. Using Gauss' law, prove (*a*) that $E = 0$ for $r < a$ and (*b*) that between the cylinders E is given by

$$E = \frac{1}{2\pi\epsilon_0}\frac{\lambda}{r}.$$

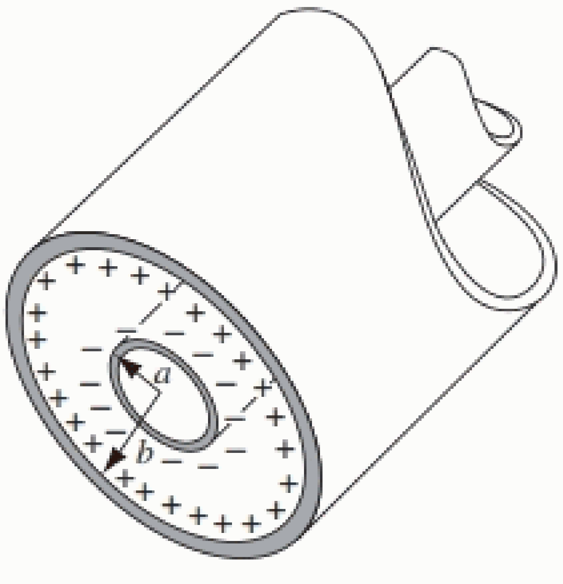


FIGURE 27-37. Problem 8.

9. In the geometry of Problem 8 a positron revolves in a circular path between and concentric with the cylinders. Find its kinetic energy, in electron-volts. Assume that $\lambda = 30$ nC/m. (Why do you not need to know the radii of the cylinders?)

10. Figure 27-38 shows a Geiger counter, used to detect ionizing radiation. The counter consists of a thin central wire, carrying positive charge, surrounded by a concentric circular conducting cylinder, carrying an equal negative charge. Thus a strong radial electric field is set up inside the cylinder. The cylinder contains a low-pressure inert gas. When a particle of radiation enters the tube through the cylinder walls, it ionizes a few of the gas atoms. The resulting free electrons are drawn to the positive wire. However, the electric field is so intense that, between collisions with the gas atoms, they have gained energy sufficient to ionize these atoms also. More free electrons are thereby created, and the process is repeated until the electrons reach the wire. The "avalanche" " of electrons is collected by the wire, generating a signal recording the passage of the incident particle of radiation. Suppose that the radius of the central wire is 25 μm, the radius of the cylinder 1.4 cm, and the length of the tube 16 cm. The electric field at the cylinder wall is 2.9×10^4 N/C. Calculate the amount of positive charge on the central wire. (Hint: See Problem 8.)

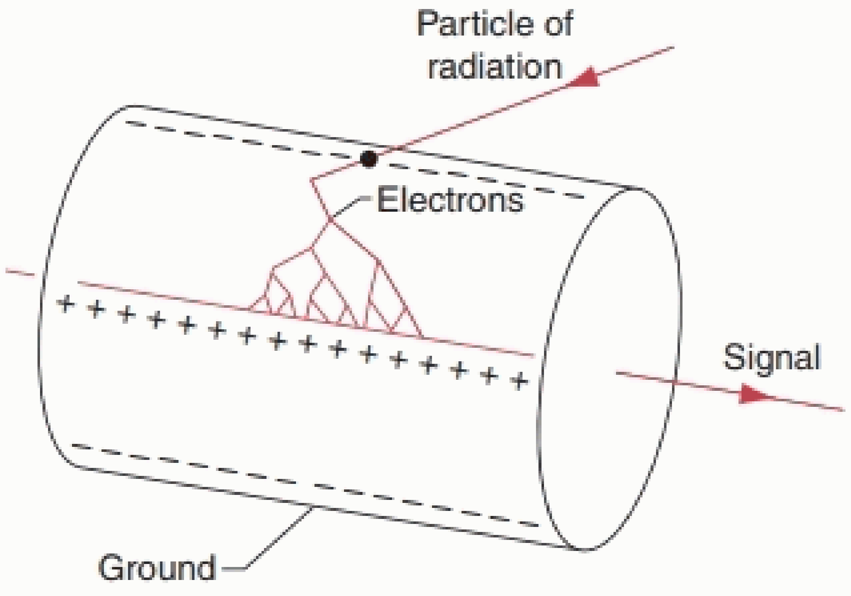


FIGURE 27-38. Problem 10.

11. An uncharged, spherical, thin, metallic shell has a point charge q at its center. Derive expressions for the electric field (*a*) inside the shell and (*b*) outside the shell, using Gauss' law. (*c*) Has the shell any effect on the field due to q? (*d*) Has the presence of q any effect on the shell? (*e*) If a second point charge is held outside the shell, does this outside charge experience a force? (*f*) Does the inside charge experience a force? (*g*) Is there a contradiction with Newton's third law here? Why or why not?

12. Charged dust particles in interstellar space, each carrying one excess electron and all of the same mass, form a stable, spherical, uniform cloud. Find the mass of each particle.

13. The spherical region $a < r < b$ carries a charge per unit volume of $\rho = A/r$, where A is a constant. At the center $(r = 0)$ of the enclosed cavity is a point charge q. What should be the value of A so that the electric field in the region $a < r < b$ has constant magnitude?

14. A spherical region carries a uniform charge per unit volume ρ. Let $\vec{\mathbf{r}}$ be the vector from the center of the sphere to a general point P within the sphere. (*a*) Show that the electric field at P is given by $\vec{\mathbf{E}} = \rho\vec{\mathbf{r}}/3\epsilon_0$. (*b*) A spherical cavity is created in the sphere, as shown in Fig. 27-39. Using superposition concepts, show that the electric field at all points within the cavity is $\vec{\mathbf{E}} = \rho\vec{\mathbf{a}}/3\epsilon_0$ (uniform field), where $\vec{\mathbf{a}}$ is the vector connecting the center of the sphere with the center of the cavity. Note that both these results are independent of the radii of the sphere and the cavity.

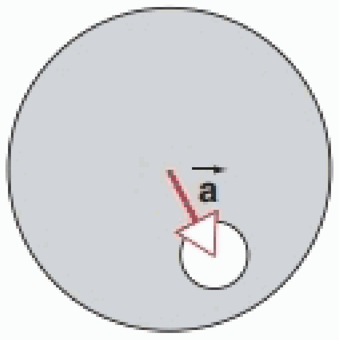


FIGURE 27-39. Problem 14.

15. Show that stable equilibrium under the action of electrostatic forces alone is impossible. (Hint: Assume that at a certain point P in an electric field $\vec{\mathbf{E}}$, a charge $+q$ would be in stable equilibrium if it were placed there. Draw a spherical Gaussian surface about P, imagine how $\vec{\mathbf{E}}$ must point on this surface, and apply Gauss' law to show that the assumption leads to a contradiction.) This result is known as Earnshaw's theorem.

16. A plane slab of thickness d has a uniform volume charge density ρ. Find the magnitude of the electric field at all points in space both (*a*) inside and (*b*) outside the slab, in terms of x, the distance measured from the median plane of the slab.

17. A solid nonconducting sphere of radius R carries a nonuniform charge distribution, with charge density $\rho = \rho_S r/R$, where ρ_S is a constant and r is the distance from the center of the sphere. Show that (*a*) the total charge on the sphere is $Q = \pi\rho_S R^3$ and (*b*) the electric field inside the sphere is given by

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{R^4}r^2.$$

18. An insulated conductor of arbitrary shape carries a net charge of $+10\ \mu C$. Inside the conductor is a hollow cavity within which is a point charge $q = +3.0\ \mu C$. What is the charge (*a*) on the cavity wall and (*b*) on the outer surface of the conductor?

19. A conducting sphere carrying charge Q is surrounded by a spherical conducting shell. (*a*) What is the net charge on the inner surface of the shell? (*b*) Another charge q is placed outside the shell. Now what is the net charge on the inner surface of the shell? (*c*) If q is moved to a position between the shell and the sphere, what is the net charge on the inner surface of the shell? (*d*) Are your answers valid if the sphere and shell are not concentric?

COMPUTER PROBLEMS

1. Verify Gauss' law with a numerical calculation. A point charge $q = 1$ nC is located 0.5 m outside the surface of a sphere of radius $r = 1.0$ m. Calculate the electric flux through the sphere.

2. Verify Gauss' law with a numerical calculation. A point charge $q = 1$ nC is located halfway between the center and the surface of a sphere of radius $r = 1.0$ m. Calculate the electric flux through the sphere.

3. A point charge of $q = 1.0\ \mu C$ is located on the axis of a cylindrical surface of radius $r = 0.5$ m and length $L = 3.0$ m. The point charge is 1.0 m from one end, and 2.0 m from the other. (*a*) Numerically calculate the electric flux through the curved portion of the cylinder. (*b*) Analytically verify your answer. (Note: No integration is required here!)

CHAPTER 28

ELECTRIC POTENTIAL ENERGY AND POTENTIAL

In Chapters 11 through 13 we learned that methods based on energy concepts offered new insights in understanding mechanics and often provided simplifications in solving mechanics problems. In Chapter 14 we used methods based on potential energy in situations involving the gravitational force to determine such properties as the motions of satellites and planets.

In this chapter we introduce the energy method to the study of electrostatics. We begin with electric potential energy, which we shall find can characterize an electrostatic force just as gravitational potential energy can characterize a gravitational force. We then generalize to the concept of electric potential and show how to find the electric potential for various discrete and continuous charge distributions.

28-1 POTENTIAL ENERGY

Many electrical phenomena are associated with the transfer of large quantities of energy. For example, when a lightning flash strikes the Earth from a cloud, an energy of typically 10^8 J is released in the form of light, sound, heat, and shock wave. Where does this energy come from, and how is it stored in clouds? To understand this question, we must consider the energy associated with electrical forces.

The electrostatic force law is very similar to the gravitational force law:

$$F = \frac{1}{4\pi\epsilon_0}\frac{|q_1||q_2|}{r^2} \qquad \text{electrostatic,} \qquad (28\text{-}1a)$$

$$F = G\frac{m_1 m_2}{r^2} \qquad \text{gravitational.} \qquad (28\text{-}1b)$$

Both forces depend on the inverse square of the separation distance between the two objects. When an object moves from place to place under the gravitational force of another object (which we assume to remain at rest), the work done by the gravitational force on the first object depends only on the starting and finishing points and does not depend on the path taken between the points. In Section 12-1 we described a force that has this special property as a *conservative force,* and we concluded in Section 12-2 that for a conservative force we could define a *potential energy.* The difference in potential energy ΔU of the system as the object moves from its initial to its final position is equal to the negative of the work done by the force:

$$\Delta U = U_\mathrm{f} - U_\mathrm{i} = -W_\mathrm{if} = -\int_\mathrm{i}^\mathrm{f} \vec{\mathbf{F}} \cdot d\vec{\mathbf{s}}, \qquad (28\text{-}2)$$

where W_if is the work done by the force $\vec{\mathbf{F}}$ when the object moves from i to f. In the case of the gravitational force, we showed in Section 14-6 that, when an object with mass m_2 moves from a distance r_i from mass m_1 to a distance r_f from m_1, the potential energy difference is

$$\Delta U = -Gm_1m_2\left(\frac{1}{r_\mathrm{f}} - \frac{1}{r_\mathrm{i}}\right). \qquad (28\text{-}3)$$

This potential energy difference is associated with the entire system consisting of m_1 and m_2, not with either object alone.

Because of the similarity of the electrostatic and gravitational force laws, we can make the same conclusion about the electrostatic force that we did about the gravitational force: *The electrostatic force is conservative, and therefore there is a potential energy associated with the configuration* *(the relative locations of the objects) of a system in which electrostatic forces act.*

Why is this approach useful for electrostatic forces? In mechanics, we learned that there are two ways to analyze problems. One approach is based on force (a vector) and allows us to determine the position and velocity of an object at every point of its motion. The other approach is based on energy (a scalar) and allows us to determine how a system changes in moving from a certain initial state to a certain final state. We will find in a similar way that both approaches are useful when we study interactions between charged objects.

There is one important property in which the electrostatic force differs from the gravitational force: the gravitational force is always attractive, whereas (depending on the relative signs of the charges) electrostatic forces can be either attractive or repulsive. This difference can affect the sign of the potential energy, but it in no way changes our argument based on the similarity of the two forces.

28-2 ELECTRIC POTENTIAL ENERGY

In this section we use the electrostatic force discussed in Chapter 25 to obtain the electric potential energy due to the interaction between two electric charges, and we extend the calculation to include the case of a collection of more than two charges.

Accepting the conclusion of the previous section that the electrostatic force is conservative, we can calculate the change in potential energy when a charge q_2 moves from point a to point b subject to the force due to another charge q_1 at rest. Let us assume for the present that both charges are positive. Figure 28-1 shows the geometry of the process. We have simplified the problem slightly by assuming that the motion from a to b is along the line that connects q_1 and q_2. (Later we shall generalize to other kinds of displacements.) We take the origin to be at the location of q_1, and we let r represent the location of q_2 relative to this origin. In Eq. 28-2, the vector $d\vec{\mathbf{s}}$ represents an infinitesimal displacement along the direction of motion from a to b. The force $\vec{\mathbf{F}}$ and the displacement $d\vec{\mathbf{s}}$ are always parallel for this motion, and so $\vec{\mathbf{F}} \cdot d\vec{\mathbf{s}} = F\,ds$. For the motion shown in Fig. 28-1, $ds = dr$ because the displacement is always in the direction of r. With these substitutions, Eq. 28-2 becomes

$$\Delta U = -\int_a^b \vec{\mathbf{F}} \cdot d\vec{\mathbf{s}} = -\int_a^b F\,dr = -\int_{r_a}^{r_b} \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2}\,dr. \tag{28-4}$$

Carrying out the integral, we obtain

$$\Delta U = U_b - U_a = \frac{1}{4\pi\epsilon_0} q_1 q_2 \left(\frac{1}{r_b} - \frac{1}{r_a}\right). \tag{28-5}$$

Equation 28-5 is valid whether the motion of q_2 is toward or away from q_1. If q_2 moves toward q_1, then $r_b < r_a$ and $\Delta U > 0$; that is, the potential energy increases if the charges move closer together. If q_2 moves away from q_1, then $r_b > r_a$ and $\Delta U < 0$; that is, the potential energy decreases if the charges move further apart.

Equation 28-5 also remains valid whether the signs of the charges are positive or negative. If both charges are negative, clearly we obtain the same result. If the charges have opposite signs (one positive and the other negative), then the force between them is attractive. With the force vector in Fig. 28-1 in the opposite direction, we have

$$\begin{aligned}\vec{\mathbf{F}} \cdot d\vec{\mathbf{s}} &= -F\,ds = -F\,dr = -\frac{1}{4\pi\epsilon_0}\frac{|q_1||q_2|}{r^2}\,dr \\ &= \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2}\,dr,\end{aligned} \tag{28-6}$$

where the last step can be made because $q_1 q_2 = -|q_1||q_2|$ when one of the charges is negative and one is positive. This gives exactly the same integrand as Eq. 28-4, and so the integral gives the same result.

When the charges have opposite signs, so that $q_1 q_2$ is negative in Eq. 28-5, then $\Delta U < 0$ when the charges move closer together and $\Delta U > 0$ when the charges move further apart.

Suppose we move q_2 in a direction that is *not* along the line connecting q_1 and q_2. Figure 28-2 shows q_2 moving from a to b along an arc of a circle of radius r centered at q_1. Along this path, $\vec{\mathbf{F}}$ is always perpendicular to $d\vec{\mathbf{s}}$, and so $\vec{\mathbf{F}} \cdot d\vec{\mathbf{s}} = 0$ throughout the path. The electrostatic force does no work along this path, and so $\Delta U = 0$.

To move q_2 between arbitrary points a and b, as in Fig. 28-3, we can choose a variety of possible paths. Along

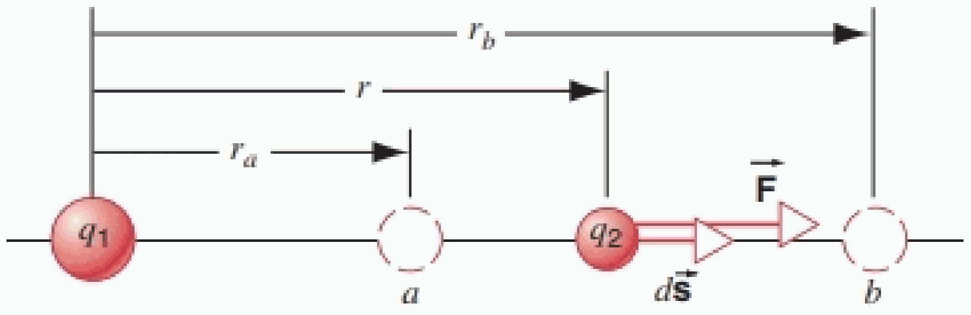


FIGURE 28-1. A charged particle q_2 moves from a to b under the influence of the electrostatic force $\vec{\mathbf{F}}$ exerted by q_1. The points a and b lie along the line connecting q_1 and q_2.

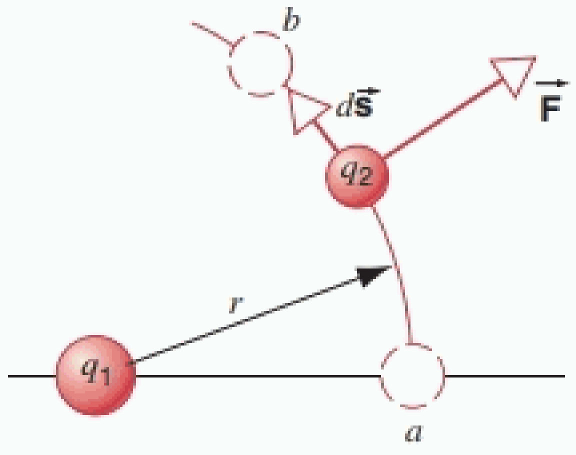


FIGURE 28-2. The motion of q_2 from a to b is now along a path of constant radius r.

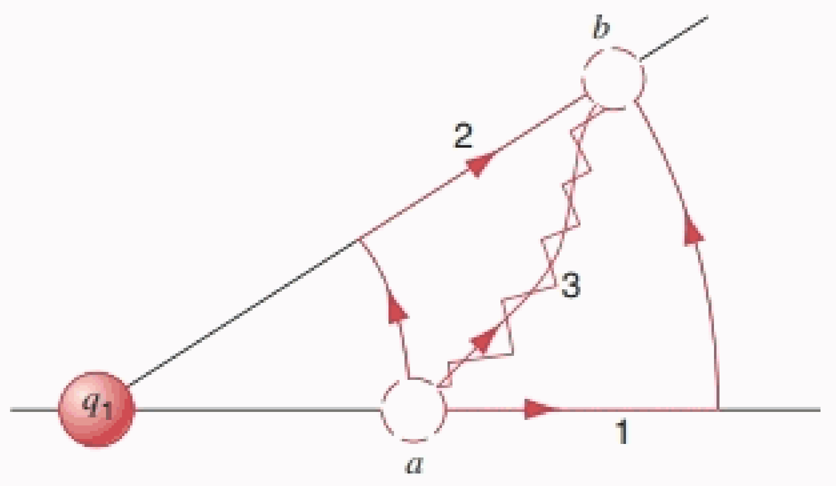


FIGURE 28-3. q_2 moves between arbitrary points *a* and *b* along several possible paths.

paths 1 and 2, ΔU is given by Eq. 28-5 for the radial (straight) parts of the paths and $\Delta U = 0$ for the tangential (curved) parts of the paths. The arbitrary path 3 can be broken into a series of short radial and tangential steps. Along each tangential step $\Delta U = 0$, while the total ΔU along all the radial steps is given by Eq. 28-5. Our conclusion is that Eq. 28-5 gives ΔU for any path between point *a*, which is a distance r_a from q_1, and point *b*, which is a distance r_b from q_1, no matter where the points may be located. This is consistent with our assertion that the electrostatic force is conservative, which means that the work and therefore the change in potential energy in moving from *a* to *b* does not depend on the path.

So far we have been discussing the *difference* in potential energy between two points: $\Delta U = U_b - U_a$. We can extend the discussion to define the potential energy at a single point *b* by choosing a reference point *a* of potential energy and assigning a reference value to the potential energy U_a at that point. Often it is convenient to choose the reference point to correspond to an infinite separation between the charges, and we generally choose the reference value $U_a = 0$. Then, letting point *b* represent any point where the separation is *r*, Eq. 28-5 becomes

$$U(r) = \frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{r}. \tag{28-7}$$

In this expression, *U* is positive whenever q_1 and q_2 have the same sign, corresponding to a repulsive force, and *U* is negative whenever q_1 and q_2 have opposite signs, corresponding to an attractive force. If we compare Eq. 28-7 with the corresponding expression given in Eq. 14-17 for the gravitational potential energy, $U(r) = -Gm_1m_2/r$, we note that the gravitational potential energy is always negative, because the gravitational force is always attractive. This agrees with the negative value of the electrostatic potential energy when the charges have opposite signs and the force is attractive.

Conservation of Energy in Electrostatics

In an isolated system of two charges, the total mechanical energy $E = K + U$ is conserved. Let us assume that q_1 is held in a fixed position and q_2 is released from rest at a certain distance from q_1. If the two charges have the same sign, then $\Delta U < 0$ as q_2 is pushed away from q_1 by the repulsive force. Conservation of total mechanical energy then requires that $\Delta K > 0$, so the speed of q_2 must increase. If instead we propel q_2 toward q_1 with a certain initial kinetic energy, then $\Delta U > 0$ as the separation decreases; conservation of energy then requires that $\Delta K < 0$, and so the speed of q_2 decreases. These conclusions are reversed if the charges have opposite signs so that the force is attractive.

Here is another way to view energy conservation in a system of two charges. Let us assume that the two charges have the same sign. We begin with the charges at rest separated by a very large distance, and we move q_2 and place it at rest a certain distance from q_1. To accomplish this task, the external agent that moves q_2 must exert a force to oppose the electrostatic repulsion between q_1 and q_2. In doing so the agent does positive external work on the system, and so the energy of the system increases. The energy of the system has increased by an amount ΔU as a result of the work done by the external agent. Put another way, the external agent has stored energy in the system, in exact analogy to the storage of energy when an external agent compresses a spring. By releasing the charges, we could recover the stored energy as kinetic energy of the moving charges.

If instead the charges have opposite signs so that the electrostatic force is attractive, then the external agent does negative external work on the system to move q_2 from a large separation to place it at rest in a location with a smaller separation from q_1. This work decreases the energy stored in the system and so is not recoverable. (Without the external agent, q_2 would on its own accelerate toward q_1; the agent must expend energy in restraining q_2 in order to place it at rest at the specified location.)

If q_1 and q_2 have opposite signs and begin with a small separation, then the external agent must do positive work equal to ΔU in order to separate the charges to a large distance. When this concept is applied to atoms or molecules, this energy may be called the *binding energy,* the *ionization energy,* or the *dissociation energy.* This quantity represents the external energy we must supply, for example, to remove an electron from an atom or to divide a molecule such as KCl into K^+ and Cl^- ions.

SAMPLE PROBLEM 28-1. Two protons in the nucleus of a ^{238}U atom are 6.0 fm apart. What is the potential energy associated with the electric force that acts between these two particles?

Solution From Eq. 28-5, with $q_1 = q_2 = +1.60 \times 10^{-19}$ C, we obtain

$$U = \frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{r} = \frac{(8.99 \times 10^9\ \mathrm{N\cdot m^2/C^2})(1.60 \times 10^{-19}\ \mathrm{C})^2}{6.0 \times 10^{-15}\ \mathrm{m}}$$

$$= 3.8 \times 10^{-14}\ \mathrm{J} = 2.4 \times 10^5\ \mathrm{eV} = 240\ \mathrm{keV},$$

where we take $U = 0$ for the configuration in which the protons are far apart. The two protons do not fly apart because they are held together by the attractive *strong force* that binds the nucleus together. Unlike the electric force, there is no simple potential energy function that represents the strong force.

SAMPLE PROBLEM 28-2. Two objects, one with mass $m_1 = 0.0022$ kg and charge $q_1 = +32\ \mu\text{C}$ and the other with mass $m_2 = 0.0039$ kg and charge $q_2 = -18\ \mu\text{C}$, are initially a distance 4.6 cm apart. With object 1 held in a fixed position, object 2 is released from rest. What is the speed of object 2 when the separation between the objects is 2.3 cm? Assume that the objects behave like point charges.

Solution As the charges move closer together, with only the electrostatic force acting, the reduction in potential energy must be balanced by a corresponding increase in the kinetic energy. Let the initial condition be the instant that object 2 is released (with $K_i = 0$) and the final condition be the instant that their separation is 2.3 cm. Then conservation of energy gives $U_i + K_i = U_f + K_f$, or (with $K_i = 0$)

$$\begin{aligned} K_f = U_i - U_f = -\Delta U &= -\frac{q_1 q_2}{4\pi\epsilon_0}\left(\frac{1}{r_f} - \frac{1}{r_i}\right) \\ &= -(8.99 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)(32 \times 10^{-6}\ \mu\text{C}) \\ &\quad \times (-18 \times 10^{-6}\ \mu\text{C})\left(\frac{1}{0.023\ \text{m}} - \frac{1}{0.046\ \text{m}}\right) \\ &= 113\ \text{J}, \end{aligned}$$

$$v_f = \sqrt{\frac{2K_f}{m_2}} = \sqrt{\frac{2(113\ \text{J})}{0.0039\ \text{kg}}} = 240\ \text{m/s}.$$

If we instead hold object 2 fixed and release object 1, when the separation reaches 2.3 cm the kinetic energy will have the same value of 113 J, because the energy is a property of the entire system. If we released both particles from rest and allowed them to fall together, at a separation of 2.3 cm the total kinetic energy of the two particles would be 113 J. We could find the velocity of each particle using conservation of momentum.

Potential Energy of a System of Charges

Suppose we have three charges (q_1, q_2, q_3) separated by infinite distances from one another. In this configuration, $U = 0$. We wish to find the potential energy of the configuration that results after the three charges are brought closer to one another.

Let us bring the first charge q_1 in from infinity and place it at rest at the location shown in Fig. 28-4*a*. Since this charge does not interact with any other charges in this process, there is no change in potential energy; we still have $U = 0$ for the system, since the charges remain separated by infinite distances. Now we bring in charge q_2 and fix it at a distance r_{12} from q_1 (Fig. 28-4*b*). The potential energy of this configuration of q_1 and q_2 (relative to $U = 0$ at infinite separation) is $q_1 q_2 / 4\pi\epsilon_0 r_{12}$. Finally, we bring in q_3 and place it a distance r_{13} from q_1 and r_{23} from q_2 (Fig. 28-4*c*). Because q_3 interacts with both q_1 and q_2, there are two additional contributions to the potential energy of this final configuration: $q_1 q_3 / 4\pi\epsilon_0 r_{13}$ (interaction of q_1 and q_3) and $q_2 q_3 / 4\pi\epsilon_0 r_{23}$ (interaction of q_2 and q_3). The total electric potential energy of the entire system is

$$U = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r_{12}} + \frac{1}{4\pi\epsilon_0}\frac{q_1 q_3}{r_{13}} + \frac{1}{4\pi\epsilon_0}\frac{q_2 q_3}{r_{23}}. \quad (28\text{-}8)$$

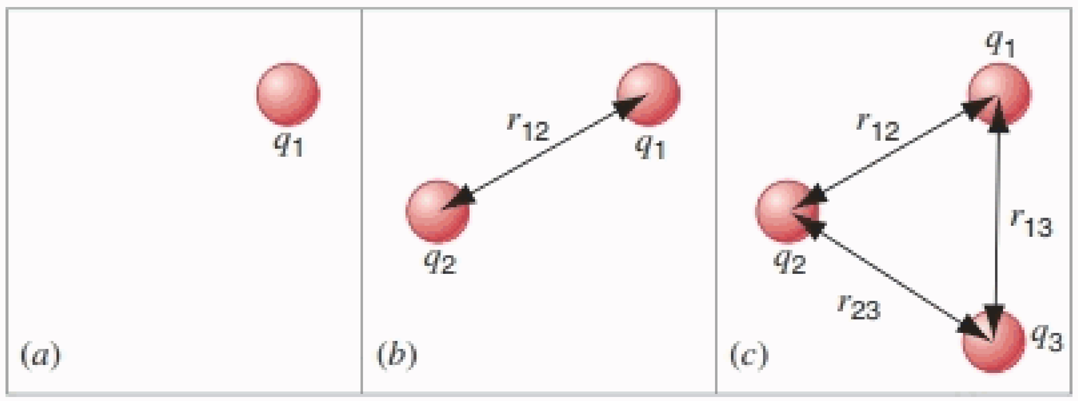


FIGURE 28-4. A system of three charges is assembled from initially infinite separations.

As Eq. 28-8 makes clear, the potential energy is a property of the system, not of any individual charge.

We could continue the process to assemble any arbitrary distribution of charges. The resulting total potential energy of any such system is independent of the order in which we assemble the charges.

From this example you can see the advantage of using an energy method to analyze this system: the sum involved in Eq. 28-8 is an *algebraic* sum of scalars. If we tried to calculate the electric field associated with a collection of three charges, we would have a more complicated *vector* sum to evaluate.

Implicit in this process is the assumption that the principle of superposition is valid. Previously we have applied this principle, which states that the interaction of any two charges is independent of the presence of other charges, to analyze vector sums. Here we see that a similar result applies to the scalar terms; for example, the potential energy term that describes the interaction of q_1 and q_3 is independent of the presence of q_2.

As we discussed above, if the external agent does positive work in assembling the charges from infinite separation (opposing a repulsive force in the process), the total potential energy calculated using Eq. 28-8 is positive. The external agent has in effect stored energy in the system of charges. If the charges are released from their positions, they will tend to fly apart, and the potential energy will decrease as the kinetic energy increases. If the total potential energy is negative, the external agent has done negative work in assembling the system of charges. In this case, the external agent must supply additional energy in the form of work to disassemble the system of charges and move them to infinite separation.

This view of potential energy can be summarized as follows:

> *The electric potential energy of a system of fixed point charges at rest is equal to the work that must be done by an external agent to assemble the system, bringing each charge in from an infinite distance where it is also at rest.*

Implicit in this view is the definition of the reference point of potential energy to be the infinite separation of the charges, where we take the reference value of the potential energy to be zero.

SAMPLE PROBLEM 28-3. In the system shown in Fig. 28-4, assume that $r_{12} = r_{13} = r_{23} = d = 12$ cm, and that

$$q_1 = +q, \quad q_2 = -4q, \quad \text{and} \quad q_3 = +2q,$$

where $q = 150$ nC. What is the potential energy of the system? Assume that $U = 0$ when the charges are infinitely far apart.

Solution Using Eq. 28-8, we obtain

$$\begin{aligned} U &= \frac{1}{4\pi\epsilon_0}\left(\frac{(+q)(-4q)}{d} + \frac{(+q)(+2q)}{d} + \frac{(-4q)(+2q)}{d}\right) \\ &= -\frac{10q^2}{4\pi\epsilon_0 d} \\ &= -\frac{(8.99 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)(10)(150 \times 10^{-9}\ \text{C})^2}{0.12\ \text{m}} \\ &= -1.7 \times 10^{-2}\ \text{J} = -17\ \text{mJ}. \end{aligned}$$

The negative potential energy in this case means that negative work would be done by an external agent to assemble this structure, starting with the three charges infinitely separated and at rest. Put another way, an external agent would have to do +17 mJ of work to dismantle the structure completely.

28-3 ELECTRIC POTENTIAL

Imagine a charge q fixed at the origin of a coordinate system. We take another charge q_0, which we call our "test charge," and we move it from r_a to r_b under the influence of the force due to q. The change in potential energy ΔU of this two-charge system is given by Eq. 28-5.

If we were to use a test charge twice as large, we would obtain twice the change in potential energy; a test charge three times as large would give three times the potential energy change.

The potential energy change is directly proportional to the size of the test charge. Put another way, the quantity $\Delta U/q_0$ is independent of the size of the test charge and is characteristic only of the central charge q. This quantity turns out to be extremely useful in analyzing a wide variety of electrostatic problems, even those that involve more complicated assemblies of charges. We define the *electric potential difference* ΔV to be the *electric potential energy difference per unit test charge:*

$$\Delta V = \frac{\Delta U}{q_0} \tag{28-9}$$

or

$$V_b - V_a = \frac{U_b - U_a}{q_0}. \tag{28-10}$$

Like the potential energy, the electric potential is a scalar. Usually we will refer to electric potential simply as "potential."

Using the relation between work and potential energy given in Eq. 28-2, we can write the definition of potential difference as

$$\Delta V = -\frac{W_{ab}}{q_0}, \tag{28-11}$$

where W_{ab} is the work done by the electrostatic force exerted by q on q_0 when the test charge moves from a to b.

By defining a suitable choice of the reference point of potential energy (such as $U_a = 0$ for an infinite initial separation of the charges), we obtained in the previous section an expression (Eq. 28-7) for the potential energy of a particular configuration rather than the change in potential energy for a change in configuration. We can do the same for electric potential. Only differences in potential have physical significance, so we are free to choose the zero point and its reference value at our convenience. When the potential is taken to be zero at points that are infinitely far from q, the electric potential is

$$V = \frac{U}{q_0}. \tag{28-12}$$

In a complicated arrangement of many charges, the potential V may be positive, negative, or zero. The potential at a point near an isolated positive charge is positive. If we were to move a positive test charge from infinity to that point, the charge would move from a location where $V = 0$ to a location where $V > 0$. Thus $\Delta V > 0$ and (according to Eq. 28-9) $\Delta U > 0$, indicating that the electric force on the test charge has done negative work. Similarly, the potential at a point near an isolated negative charge is negative; the electric force does positive work when we move a positive test charge from infinity to that point.

If the potential is zero at a point, no net work is done by the electric force as the test charge moves in from infinity to that point, although the test charge may pass through regions where it experiences attractive or repulsive electric forces. *A potential of zero at a point does not necessarily mean that the electric force is zero at that point.*

The SI unit of potential that follows from Eq. 28-9 is the joule per coulomb. This combination is given the name of *volt* (V):

$$1 \text{ volt} = 1 \text{ joule/coulomb}. \tag{28-13}$$

The common name of "voltage" is often used for the potential at a point, and we often speak of "voltage difference" instead of potential difference. When you touch the two probes of a voltmeter to two points in an electric circuit, you are measuring the voltage difference or potential difference (in volts) between those points.

We have already discussed that the electric force is conservative, and so the potential energy difference when a test charge is moved between any two points depends only on the locations of the points and not on the path taken to move from one point to the other. Equation 28-9 therefore suggests that the potential difference is similarly path independent: the potential difference between any two points in an electric field is independent of the path through which the test charge moves in traveling from one point to the other.

For any arbitrary potential difference ΔV, no matter what the arrangement of charges that produces it, we can write Eq. 28-9 as

$$\Delta U = q\,\Delta V. \qquad (28\text{-}14)$$

This equation indicates that when *any* charge q moves between two points whose potential difference is ΔV, the system experiences a change in potential energy ΔU given by Eq. 28-14. The potential difference ΔV is set up by other charges that are fixed at rest, so that the motion of q does not change ΔV. In using Eq. 28-14, we see from Eq. 28-13 that if ΔV is expressed in volts and q is in coulombs, then ΔU comes out in joules.

From Eq. 28-14, you can see that the *electron-volt,* which we have introduced previously as a unit of energy, follows directly from the definition of potential or potential difference. If ΔV is expressed in volts and q in units of the elementary charge e, then ΔU is expressed in electron-volts (eV). For example, consider a system in which a carbon atom from which all six electrons have been removed ($q = +6e$) moves through a change in potential of $\Delta V = +20$ kV. The change in potential energy is

$$\Delta U = q\,\Delta V = (+6e)(+20\text{ kV}) = +120\text{ keV}.$$

Doing such calculations in units of eV is a great convenience when dealing with atoms or nuclei, in which the charge is easily expressed in terms of e.

Keep in mind that *potential differences* are of fundamental concern and that Eq. 28-12 depends on the arbitrary assignment of the value zero to the potential at the reference position (infinity); this reference potential could equally well have been chosen as any other value—say, -100 V. Similarly, any other agreed-upon point could be chosen as a reference position. In many problems the Earth is taken as a reference of potential and is assigned the value zero. The location of the reference point and the value of the potential there are chosen for convenience; other choices would change the potential everywhere by the same amount but would not change the potential difference between any two points.

SAMPLE PROBLEM 28-4. An alpha particle ($q = +2e$) in a nuclear accelerator moves from one terminal at a potential of $V_a = +6.5 \times 10^6$ V to another at a potential of $V_b = 0$. (*a*) What is the corresponding change in the potential energy of the system? (*b*) Assuming that the terminals and their charges do not move and that no external forces act on the system, what is the change in kinetic energy of the particle?

Solution (*a*) From Eq. 28-14, we have

$$\begin{aligned}\Delta U = U_b - U_a &= q(V_b - V_a)\\ &= (+2)(1.6 \times 10^{-19}\text{ C})(0 - 6.5 \times 10^6\text{ V})\\ &= -2.1 \times 10^{-12}\text{ J}.\end{aligned}$$

(*b*) If no external force acts on the system, then its mechanical energy $E = U + K$ must remain constant. That is, $\Delta E = \Delta U + \Delta K = 0$, and so

$$\Delta K = -\Delta U = +2.1 \times 10^{-12}\text{ J}.$$

The alpha particle gains a kinetic energy of 2.1×10^{-12} J, in the same way that a particle falling in the Earth's gravitational field gains kinetic energy.

To see the simplifications that result, try working this problem again with the energies expressed in units of eV.

28-4 CALCULATING THE POTENTIAL FROM THE FIELD

So far we have characterized electric charges and their interactions using four different properties: electric force, electric field, electric potential energy, and electric potential. Table 28-1 shows these four properties. Two of them are vectors (force and field), and two are scalars (potential energy and potential). Two of them characterize the interactions of two particles with one another (force and potential energy), and two of them represent the effect at a point in space due to a single charge or collection of charges (field and potential). The double arrows in the table show that the quantities in adjacent boxes of the table can be calculated from one another; for example, $\vec{\mathbf{E}}$ from $\vec{\mathbf{F}}$ (Eq. 26-3), U from $\vec{\mathbf{F}}$ (Eq. 28-4), and V from U (Eq. 28-12). Now we examine the fourth connection—namely, that between V and $\vec{\mathbf{E}}$.

The connection between V and $\vec{\mathbf{E}}$ follows directly from the definition of potential in Eq. 28-11: $\Delta V = -W_{ab}/q_0$. Suppose we move a test charge q_0 from a to b in an electric field $\vec{\mathbf{E}}$. Calculating the work done by the electric force $\vec{\mathbf{F}} = q_0\vec{\mathbf{E}}$, we obtain

$$\Delta V = \frac{-W_{ab}}{q_0} = \frac{-\int_a^b \vec{\mathbf{F}} \cdot d\vec{\mathbf{s}}}{q_0} = \frac{-\int_a^b q_0\vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}}{q_0}$$

or

$$\Delta V = V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}. \qquad (28\text{-}15)$$

If the electric field is along the direction of $d\vec{\mathbf{s}}$, then the integral in Eq. 28-15 will be positive, and the potential difference will be negative; that is, $V_b < V_a$. The electric field

TABLE 28-1 Properties of Electric Charges

	Vector Description	*Scalar Description*
Interaction between two charges	Force $\vec{\mathbf{F}}$ ⟷	Potential energy U
	↕	↕
Effect of one charge or group of charges at a point in space	Field $\vec{\mathbf{E}}$ ⟷	Potential V

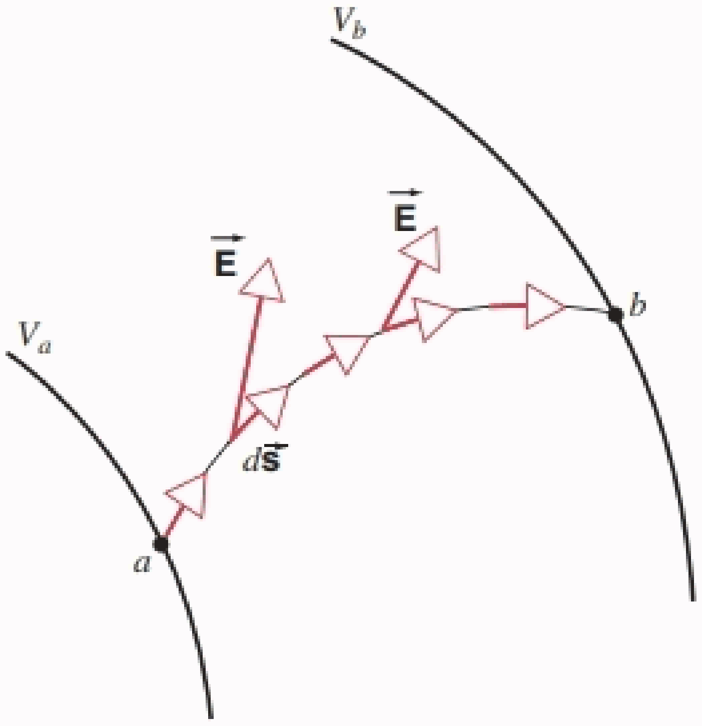


FIGURE 28-5. The potential difference between a and b can be found by calculating the line integral of $\vec{\mathbf{E}}$ along the path ab.

would move a positively charged particle from a region of higher potential to a region of lower potential or a negatively charged particle in the opposite direction.

An integral of the form of Eq. 28-15 is called a *line integral.* Figure 28-5 illustrates the calculation of a line integral. We integrate from a to b along any convenient path; we know that the potential difference is a path-independent quantity, so we get the same result from Eq. 28-15 no matter what path we choose. In general the magnitude and direction of $\vec{\mathbf{E}}$ may change from point to point along the path. At each step of the path, we find the dot product between $\vec{\mathbf{E}}$ and the path increment $d\vec{\mathbf{s}}$ (which essentially gives the component of $\vec{\mathbf{E}}$ along the path), and we add up these dot products for the entire path.

As we did in Section 28-3, we may wish to find the potential at a point, relative to some chosen reference potential, rather than the potential difference given by Eq. 28-15. If we choose the reference point to be at infinity and define $V = 0$ as the reference, then Eq. 28-15 gives for the potential at point P

$$V_P = -\int_{\infty}^{P} \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}. \qquad (28\text{-}16)$$

SAMPLE PROBLEM 28-5. In Fig. 28-6, a test charge q_0 moves through a uniform electric field $\vec{\mathbf{E}}$ from a to b along the path acb. Find the potential difference between a and b.

Solution For the path ac we have, from Eq. 28-15,

$$V_c - V_a = -\int_a^c \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -\int_a^c E\,ds\cos(\pi - \theta)$$

$$= E\cos\theta \int_a^c ds.$$

The integral is the length of the line ac, which is $L/\cos\theta$. Thus

$$V_c - V_a = E\cos\theta \frac{L}{\cos\theta} = EL.$$

Points b and c have the same potential, because no work is done in moving a charge between them, $\vec{\mathbf{E}}$ and $d\vec{\mathbf{s}}$ being at right angles for all points on the line cb. Thus

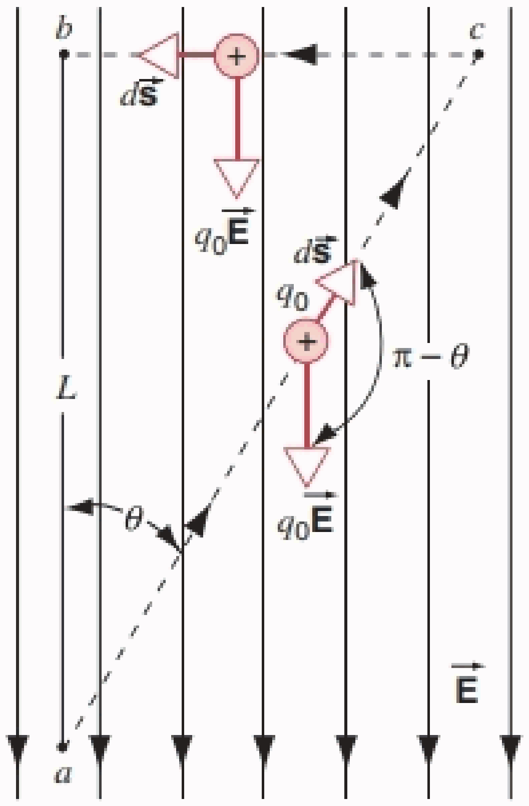


FIGURE 28-6. Sample Problem 28-5. A test charge q_0 moves along the path acb through the uniform electric field $\vec{\mathbf{E}}$.

$$V_b - V_a = (V_b - V_c) + (V_c - V_a) = 0 + EL = EL.$$

This is the same value derived for a direct path connecting a and b, a result to be expected because the potential difference between two points is independent of path.

28-5 POTENTIAL DUE TO POINT CHARGES

In this section we will use the results of the previous sections to obtain the potential for various arrangements of point charges. In the next section we discuss the potential due to continuous charge distributions.

We first consider the potential due to a positive point charge q. Let a test charge q_0 move from point a to point b in the vicinity of q. We wish to use the test charge to find the potential difference between points a and b due to q. We can use the geometry of Fig. 28-1, with q_1 replaced by q and q_2 replaced by q_0.

We have already found the potential energy difference ΔU for this situation, which was given by Eq. 28-5 for two point charges. Writing Eq. 28-5 for charges q and q_0, and using Eq. 28-9 for the potential difference, we find

$$V_b - V_a = \frac{U_b - U_a}{q_0} = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{r_b} - \frac{1}{r_a}\right). \qquad (28\text{-}17)$$

As we discussed in Section 28-2, Eq. 28-5 holds even if points a and b do not lie on the same line. Equation 28-17 is valid for the potential difference between any two points a and b.

Instead of the potential difference between two points, we can find the potential at a single point in the vicinity of q. Equation 28-7 gives the potential energy U due to the interaction of two point charges. The reference point for this expression is taken at infinity, where we define $U = 0$. We

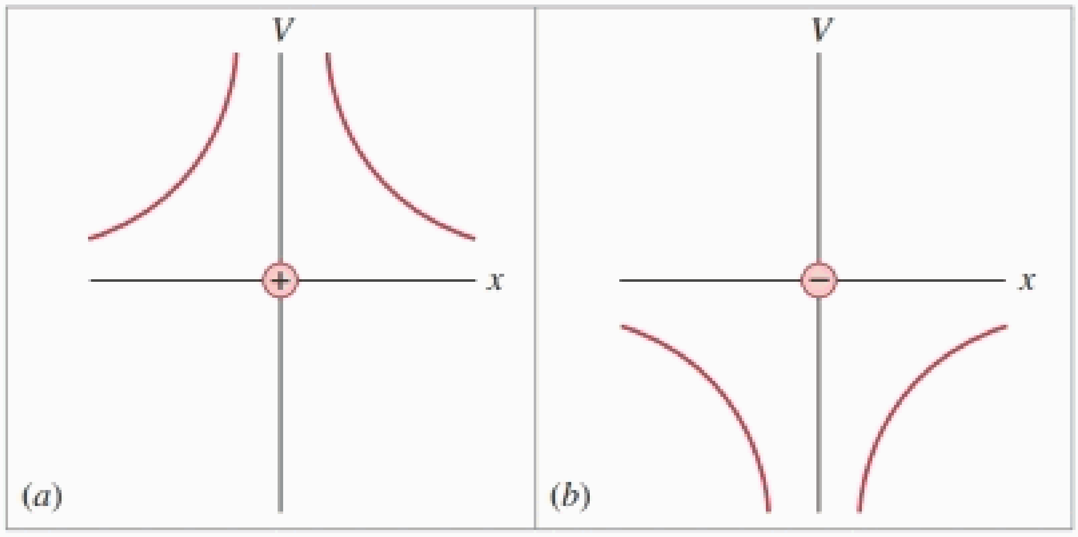


FIGURE 28-7. The potential along one dimension (chosen to be the x axis) for (a) a positive point charge and (b) a negative point charge. The magnitude of the potential increases to infinity as the distance from the charge becomes zero. The potential for a single positive charge is positive everywhere, and for a single negative charge the potential is negative everywhere.

can use Eq. 28-7, written for charges q and q_0, to find the potential at a point using Eq. 28-12 for the potential:

$$V = \frac{U}{q_0} = \frac{1}{4\pi\epsilon_0}\frac{q}{r}, \qquad (28\text{-}18)$$

for any point at a distance r from q. Note that Eq. 28-18 could have been obtained directly from Eq. 28-17 by imposing the reference condition with $V_a = 0$ at $r_a = \infty$.

Equation 28-18 shows that the potential for a single positive point charge is zero at large distances and grows to large positive values as we approach the charge ($r \to 0$). If q is negative, the potential grows to large negative values as we approach the charge. Note that these results do not depend at all on the sign of the test charge q_0 we use in the calculation. Figure 28-7 shows the potential as a function of the distance from the charge for a positive and a negative point charge.

SAMPLE PROBLEM 28-6. What must be the magnitude of an isolated positive point charge for the electric potential at 15 cm from the charge to be +120 V? Assume that $V = 0$ at infinity.

Solution Solving Eq. 28-18 for q yields

$$\begin{aligned} q = 4\pi\epsilon_0 rV &= (4\pi)(8.9 \times 10^{-12}\ \mathrm{C^2/N\cdot m^2})(0.15\ \mathrm{m})(120\ \mathrm{V}) \\ &= 2.0 \times 10^{-9}\ \mathrm{C} = 2.0\ \mathrm{nC}. \end{aligned}$$

This charge is comparable to charges that can be produced by friction, such as by rubbing a balloon.

SAMPLE PROBLEM 28-7. What is the electric potential at the surface of a gold nucleus? The radius is 7.0×10^{-15} m, and the atomic number Z is 79.

Solution The nucleus, assumed to be spherically symmetric, behaves electrically for external points as if it were a point charge. Thus we can use Eq. 28-18, which gives, with $q = +79e$,

$$\begin{aligned} V = \frac{1}{4\pi\epsilon_0}\frac{q}{r} &= \frac{(9.0 \times 10^9\ \mathrm{N\cdot m^2/C^2})(79)(1.6 \times 10^{-19}\ \mathrm{C})}{7.0 \times 10^{-15}\ \mathrm{m}} \\ &= 1.6 \times 10^7\ \mathrm{V}. \end{aligned}$$

This large positive potential has no effect outside a gold *atom* because it is compensated by an equally large negative potential from the 79 atomic electrons of gold.

Potential Due to a Collection of Point Charges

Suppose we have a collection of N point charges q_1, q_2, . . . , q_N located at various fixed points (Fig. 28-8). We wish to find the potential at an arbitrary point P due to this collection of charges. The procedure is to calculate the potential at P due to each charge as if the others were not present and then add the resulting potentials to get the total potential. That is,

$$\begin{aligned} V &= V_1 + V_2 + \cdots + V_N \\ &= \frac{1}{4\pi\epsilon_0}\frac{q_1}{r_1} + \frac{1}{4\pi\epsilon_0}\frac{q_2}{r_2} + \cdots + \frac{1}{4\pi\epsilon_0}\frac{q_N}{r_N}, \end{aligned} \qquad (28\text{-}19)$$

which we can write in compact form as

$$V = \frac{1}{4\pi\epsilon_0}\sum_{n=1}^{N}\frac{q_n}{r_n}. \qquad (28\text{-}20)$$

In these expressions, q_n is the value (magnitude and sign) of the nth charge and r_n is the distance of the nth charge from the point P where we wish to find the potential.

We could use Eq. 28-20, for instance, to find the work done when we bring a test charge q_0 in from infinity to point P in Fig. 28-8. For this calculation we see the advantage of using the potential, which is a scalar, rather than the force, which is a vector. To find the net force on a test charge at P, it would be necessary to find a vector sum. The scalar calculation of the potential is much simpler.

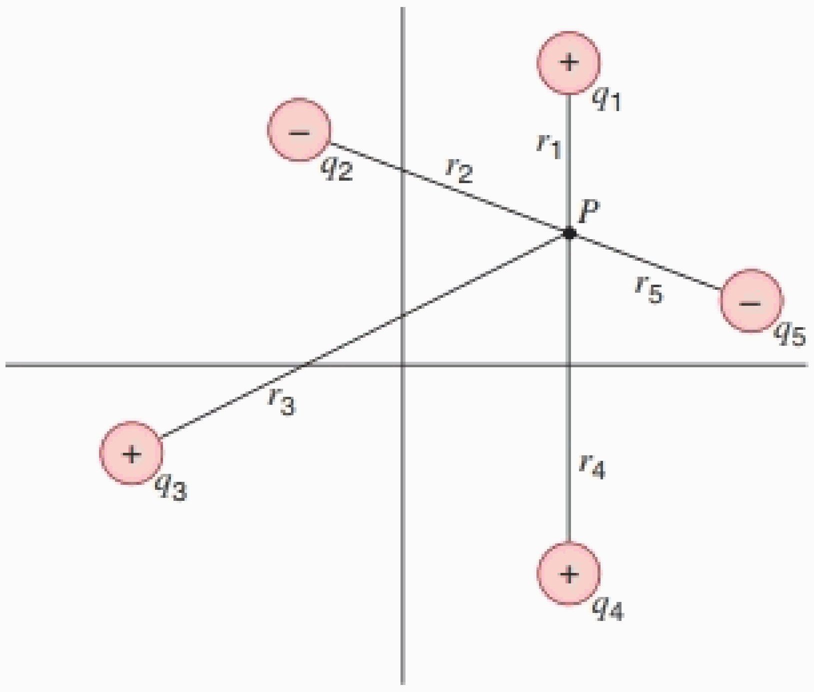


FIGURE 28-8. A collection of point charges.

In this calculation we found the contribution to the potential from each charge as if the others were not present. This is another example of the application of the principle of superposition, which we discussed in connection with electric forces in Chapter 25.

SAMPLE PROBLEM 28-8. Calculate the potential at point P, located at the center of the square of point charges shown in Fig. 28-9a. Assume that $d = 1.3$ m and that the charges are

$$q_1 = +12\text{ nC},\qquad q_3 = +31\text{ nC},$$
$$q_2 = -24\text{ nC},\qquad q_4 = +17\text{ nC}.$$

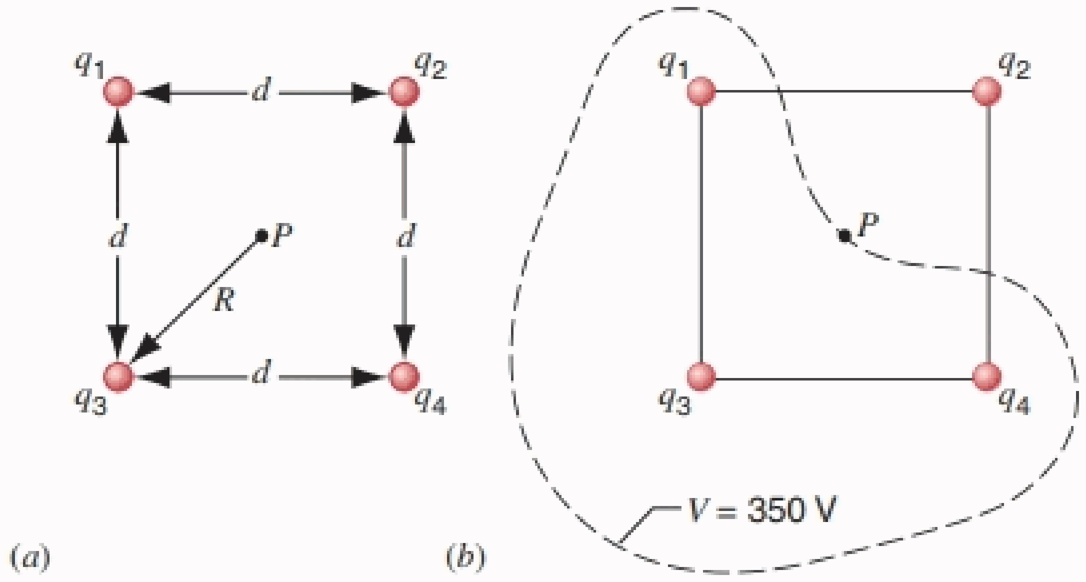
Solution From Eq. 28-20 we have

$$V = \frac{1}{4\pi\epsilon_0}\sum_n \frac{q_n}{r_n} = \frac{1}{4\pi\epsilon_0}\frac{q_1+q_2+q_3+q_4}{R}.$$

The distance R of each charge from the center of the square is $d/\sqrt{2}$ or 0.919 m, so that

$$V = \frac{(8.99\times 10^9\text{ N}\cdot\text{m}^2/\text{C}^2)(12 - 24 + 31 + 17)\times 10^{-9}\text{ C}}{0.919\text{ m}}$$
$$= 3.5\times 10^2\text{ V}.$$

Close to any of the three positive charges in Fig. 28-9a, the potential can have very large positive values. Close to the single negative charge in that figure, the potential can have large negative values. There must then be other points within the boundaries of the square that have the same potential as that at point P. The dashed line in Fig. 28-9b connects other points in the plane that have this same value of the potential. As we discuss later in Section 28-8, such *equipotential surfaces* provide a useful way of visualizing the potentials of various charge distributions.



(*a*) (*b*)

FIGURE 28-9. Sample Problem 28-8. (*a*) Four charges are held at the corners of a square. (*b*) The curve connects points that have the same potential (350 V) as the point P at the center of the square.

Potential Due to an Electric Dipole

The potential due to a electric dipole can be calculated in a straightforward way using Eq. 28-20. Figure 28-10 shows the geometry for the calculation. We place the origin of our coordinate system at the center of the dipole, and we seek the electric potential at the point P, which is located a distance r from the center of the dipole and at an angle θ from the axis of the dipole (the z axis). The distances from the positive and negative charges to P are respectively r_+ and r_-. Using Eq. 28-20, we find the potential to be

$$V = \frac{1}{4\pi\epsilon_0}\left(\frac{q}{r_+} + \frac{-q}{r_-}\right). \tag{28-21}$$

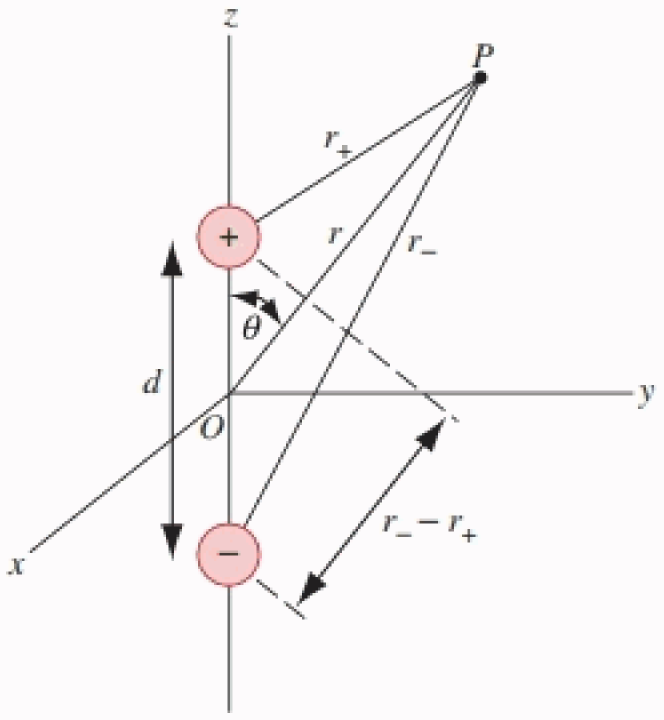


FIGURE 28-10. The geometry for calculating the potential at point P due to an electric dipole.

Equation 28-21 is the exact expression for the potential due to a dipole. However, in many applications (such as for atomic or molecular dipoles) we can obtain a more useful relationship by recognizing that our observation point P is usually very far from the dipole, compared with the distance d between the charges; that is, $r \gg d$. In this case,

$$r_- - r_+ \approx d\cos\theta \quad\text{and}\quad r_-r_+ \approx r^2,$$

and substituting these results into Eq. 28-21 we obtain

$$V = \frac{1}{4\pi\epsilon_0}\frac{qd\cos\theta}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{p\cos\theta}{r^2}, \tag{28-22}$$

where we have used Eq. 26-8 ($p = qd$) for the dipole moment. Equation 28-22 gives the potential due to a dipole at any point in space. The dipole has cylindrical symmetry for rotations about the z axis, so Eq. 28-22 is valid at points that do not lie in the plane of the diagram of Fig. 28-10.

Note that the potential due to the dipole varies as $1/r^2$. This is in contrast to the potential for a single charge, which varies (see Eq. 28-18) as $1/r$.

Equation 28-22 shows that $V = 0$ when $\theta = 90°$, which corresponds to points in the xy plane in Fig. 28-10. This means that if we move a test charge from infinity to a point in the xy plane, the dipole does no net work on the test charge. For a given r, the potential varies from positive values on the positive z axis ($\theta = 0$) to zero in the xy plane ($\theta = 90°$) to negative values on the negative z axis ($\theta = 180°$).

Note that, even though $V = 0$ in the xy plane, it is *not* true that $\vec{\mathbf{E}} = 0$ in that plane. In general we *cannot* assume that $V = 0$ implies $\vec{\mathbf{E}} = 0$ or that $\vec{\mathbf{E}} = 0$ implies $V = 0$.

SAMPLE PROBLEM 28-9. An *electric quadrupole* consists of two electric dipoles so arranged that they almost, but not quite, cancel each other in their electric effects at distant points (see Fig. 28-11). Calculate $V(r)$ for points on the axis of this quadrupole.

Solution Applying Eq. 28-20 to Fig. 28-11 yields

$$V = \frac{1}{4\pi\epsilon_0}\left(\frac{q}{r-d} + \frac{-2q}{r} + \frac{q}{r+d}\right)$$
$$= \frac{1}{4\pi\epsilon_0}\frac{2qd^2}{r(r^2-d^2)} = \frac{1}{4\pi\epsilon_0}\frac{2qd^2}{r^3(1-d^2/r^2)}.$$

Because $d \ll r$, we can neglect d^2/r^2 compared with 1, in which case the potential becomes

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^3}, \tag{28-23}$$

where Q $(= 2qd^2)$ is the *electric quadrupole moment* of the charge assembly of Fig. 28-11. Note that V varies (1) as $1/r$ for a point charge (see Eq. 28-18), (2) as $1/r^2$ for a dipole (see Eq. 28-22), and (3) as $1/r^3$ for a quadrupole (see Eq. 28-23).

Note too that (1) a dipole is two equal and opposite charges that do not quite coincide in space so that their electric effects at distant points do not quite cancel, and (2) a quadrupole is two equal and opposite dipoles that do not quite coincide in space so that their electric effects at distant points again do not quite cancel. We can continue to construct more complex assemblies of electric charges. This process turns out to be useful, because the electric potential of *any* charge distribution can be represented as a series of terms in increasing powers of $1/r$. The $1/r$ part, called the *monopole* term, depends on the net charge of the distribution, and the succeeding terms ($1/r^2$, the *dipole* term; $1/r^3$, the *quadrupole* term; and so on) indicate how the charge is distributed. This type of analysis is called an *expansion in multipoles*.

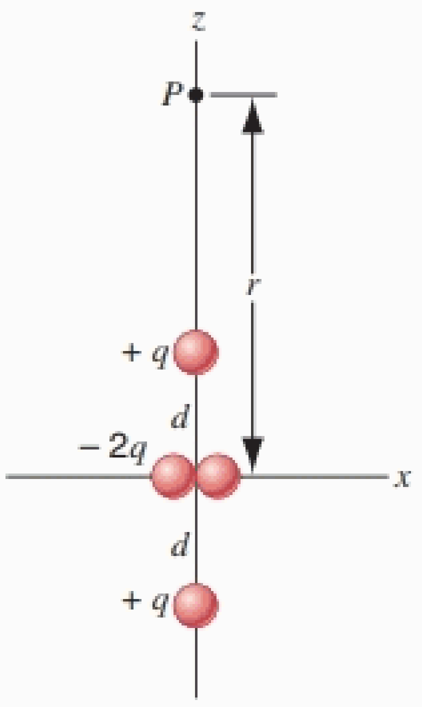


FIGURE 28-11. Sample Problem 28-9. An electric quadrupole, consisting of two oppositely directed electric dipoles.

28-6 ELECTRIC POTENTIAL OF CONTINUOUS CHARGE DISTRIBUTIONS

In Section 25-5 we introduced a procedure for calculating the force exerted by a continuous charge distribution on a point charge. We can similarly obtain the potential energy for the interaction between a continuous distribution and a point charge by calculating the potential due to the charge distribution. In this section we calculate the potential for the same three charge distributions considered in Section 25-5.

The procedure for calculating the potential for a continuous charge distribution is similar to that used to find the force (or the electric field in Section 26-4), with one important exception: the potential is a scalar, and thus we do not encounter the difficulties that arose in Section 25-5 due to the differing directions of the force elements $d\vec{\mathbf{F}}$ or field elements $d\vec{\mathbf{E}}$ from different charge elements dq.

The procedure for calculating the potential begins by dividing the object into charge elements dq. We can write the potential dV due to a charge element dq by assuming that it behaves like a point charge:

$$dV = \frac{1}{4\pi\epsilon_0}\frac{dq}{r}, \tag{28-24}$$

where r is the distance from dq to the observation point P. The total potential is found by adding the contributions from all the charge elements of the object:

$$V = \int dV = \frac{1}{4\pi\epsilon_0}\int\frac{dq}{r}, \tag{28-25}$$

where the integral is carried out over the entire charge distribution.

A Uniform Line of Charge

We can use the geometry of Fig. 28-12 to find the potential due to a uniform line of positive charge at point P, a distance y from the rod on its perpendicular bisector. Applying Eq. 28-24, using the charge element $dq = \lambda\,dz$ (where λ is the linear charge density), we have

$$dV = \frac{1}{4\pi\epsilon_0}\frac{dq}{r} = \frac{1}{4\pi\epsilon_0}\frac{\lambda\,dz}{\sqrt{z^2+y^2}}. \tag{28-26}$$

Carrying out the integration over the length L as in Eq. 28-25 and noting that y is a constant, we obtain

$$V = \frac{1}{4\pi\epsilon_0}\int_{-L/2}^{+L/2}\frac{\lambda\,dz}{\sqrt{z^2+y^2}}$$
$$= \frac{\lambda}{4\pi\epsilon_0}\left[\ln\left(z+\sqrt{z^2+y^2}\right)\right]_{-L/2}^{+L/2}$$
$$= \frac{\lambda}{4\pi\epsilon_0}\ln\left[\frac{L/2+\sqrt{L^2/4+y^2}}{-L/2+\sqrt{L^2/4+y^2}}\right], \tag{28-27}$$

where we have used the relation $\ln A - \ln B = \ln(A/B)$ to obtain the last result.

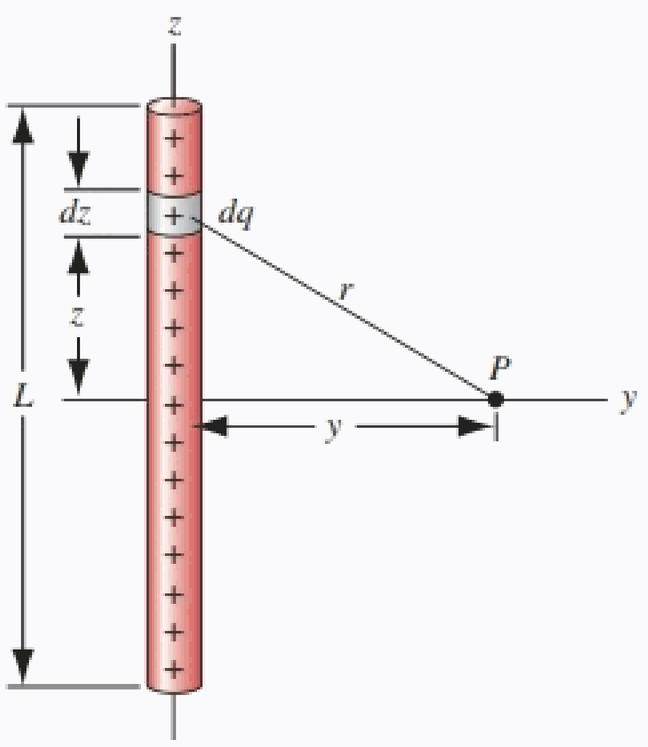


FIGURE 28-12. A uniformly charged rod. To find the potential at point P, we consider the rod to consist of many individual charge elements such as dq.

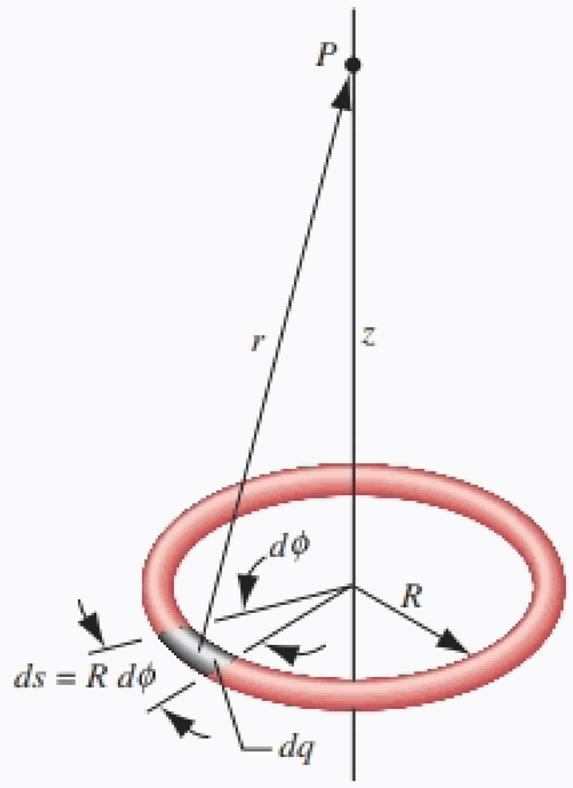


FIGURE 28-13. A uniformly charged ring. To find the potential at P, we calculate the total effect of all charge elements such as dq.

It is important to check this result to see whether it has the correct limiting value. As we move far from the rod, we expect the potential to approach 0, and Eq. 28-27 does have this property as $y \to \infty$. Furthermore, we can show that when y is large, Eq. 28-27 becomes

$$V \approx \frac{1}{4\pi\epsilon_0}\frac{\lambda L}{y} = \frac{1}{4\pi\epsilon_0}\frac{q}{y}, \tag{28-28}$$

which is simply the expression for the potential a distance y from a point charge. When we are very far away from the rod, it looks like a point charge.

A Ring of Charge

Figure 28-13 shows a uniform ring of positive charge. The contribution to the potential at point P on its axis due to the charge element $dq = \lambda\, ds = \lambda R\, d\phi$ is

$$dV = \frac{1}{4\pi\epsilon_0}\frac{dq}{r} = \frac{1}{4\pi\epsilon_0}\frac{\lambda R\, d\phi}{\sqrt{R^2 + z^2}}. \tag{28-29}$$

Integrating around the ring as we did in Section 25-5, we note that R and z both remain constant. The variable of integration is ϕ, which ranges from 0 to 2π.

$$V = \frac{1}{4\pi\epsilon_0}\frac{\lambda R}{\sqrt{R^2 + z^2}}\int_0^{2\pi} d\phi = \frac{1}{4\pi\epsilon_0}\frac{2\pi\lambda R}{\sqrt{R^2 + z^2}}. \tag{28-30}$$

Note that as $z \to \infty$, the potential decreases to zero and for large z has the approximate value $q/4\pi\epsilon_0 z$ (where $q = 2\pi\lambda R$), as expected for a location a distance z from a point charge.

A Charged Disk

With the geometry of Fig. 28-14, we can use Eq. 28-30 to find the potential dV at point P due to the ring of radius w and charge $dq = \sigma\, dA$ with area element $dA = 2\pi w\, dw$:

$$dV = \frac{1}{4\pi\epsilon_0}\frac{dq}{\sqrt{w^2 + z^2}} = \frac{1}{4\pi\epsilon_0}\frac{2\pi\sigma w\, dw}{\sqrt{w^2 + z^2}}. \tag{28-31}$$

To sum the contributions from all the rings on the disk, we integrate as w ranges from 0 to R:

$$V = \frac{\sigma}{2\epsilon_0}\int_0^R \frac{w\, dw}{\sqrt{w^2 + z^2}} = \frac{\sigma}{2\epsilon_0}\left(\sqrt{R^2 + z^2} - |z|\right). \tag{28-32}$$

The last term in Eq. 28-32 comes from evaluating $\sqrt{z^2}$ and is written as $|z|$ so that Eq. 28-32 remains valid for points on the z axis above the disk ($z > 0$) as well as below the disk ($z < 0$). The potential has its maximum value at the surface of the disk (where $z = 0$) and decreases as we move along the z axis in either direction.

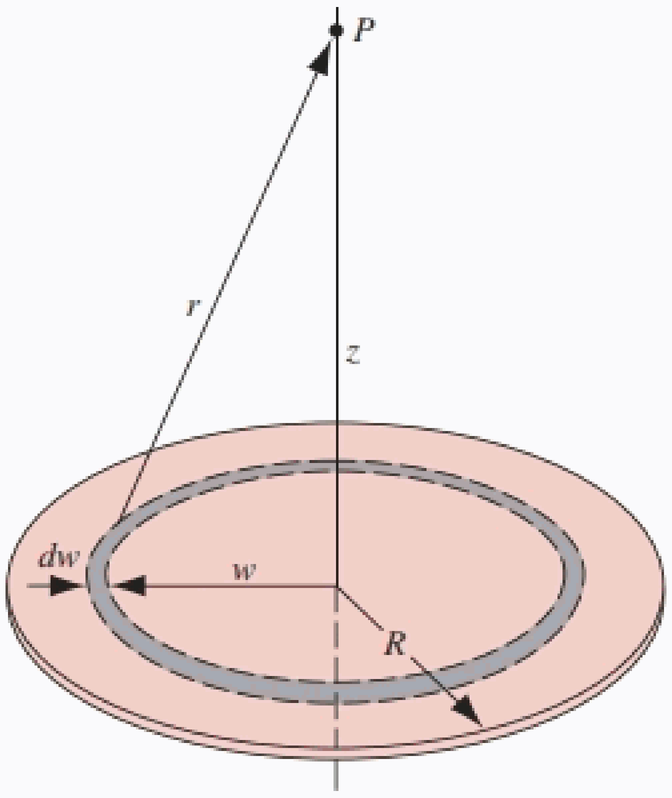


FIGURE 28-14. A disk of radius R carries a uniform charge density σ. The element of charge dq is a uniformly charged ring.

When z is large, we can use the binomial theorem to expand the square root in Eq. 28-32:

$$\sqrt{R^2+z^2} = |z|\left(1+\frac{R^2}{z^2}\right)^{1/2} \approx |z|\left(1+\frac{1}{2}\frac{R^2}{z^2}\right) \qquad (28\text{-}33)$$

and, inserting this result into Eq. 28-32, we find once again the expression for the potential of a point charge.

For very small values of z, the potential is

$$V = \sigma R/2\epsilon_0 - \sigma|z|/2\epsilon_0. \qquad (28\text{-}34)$$

The potential approaches the constant value $\sigma R/2\epsilon_0$ as $z \rightarrow 0$ and decreases linearly as z increases in either direction. The rate at which the potential decreases as we move along the axis (which is given by the second term of Eq. 28-34) is independent of the size of the disk, for a given charge density. In fact, this term turns out to be the same for any large, flat, uniformly charged plate, no matter what its size or shape (round, square, etc.), as long as we are near its center and thus far from any edge. We will use this fact in drawing a "map" of the potential in the next section.

SAMPLE PROBLEM 28-10. A disk of radius $R = 4.8$ cm carries a total charge $q = +2.5$ nC that is uniformly distributed over its surface and held in fixed locations (consider the surface to behave like an insulator). An electron is initially at rest a distance of $d = 3.0$ cm from the disk along its axis. When the electron is released, it is attracted toward the disk. What is the speed of the electron when it strikes the center of the disk?

Solution The charge density on the disk is

$$\sigma = \frac{q}{\pi R^2} = \frac{2.5\times10^{-9}\ \text{C}}{\pi(0.048\ \text{m})^2} = 3.45\times10^{-7}\ \text{C/m}^2.$$

The difference in potential between the locations with $z = d$ and $z = 0$ can be found from Eq. 28-32:

$$\begin{aligned}\Delta V = V(0) - V(d) &= \frac{\sigma R}{2\epsilon_0} - \frac{\sigma}{2\epsilon_0}\left(\sqrt{R^2+d^2} - d\right)\\ &= \frac{3.45\times10^{-7}\ \text{C/m}^2}{2(8.85\times10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2)}[0.048\ \text{m}\\ &\quad - (\sqrt{(0.048\ \text{m})^2+(0.030\ \text{m})^2} - 0.030\ \text{m})]\\ &= 417\ \text{V}.\end{aligned}$$

The change in the electron's potential energy is, from Eq. 28-14,

$$\Delta U = q\,\Delta V = (-1.60\times10^{-19}\ \text{C})(417\ \text{V}) = -6.67\times10^{-17}\ \text{J}.$$

Conservation of energy gives $\Delta U + \Delta K = 0$, or $\Delta K = -\Delta U = +6.67\times10^{-17}$ J, so

$$v = \sqrt{\frac{2K}{m}} = \sqrt{\frac{2(6.67\times10^{-17}\ \text{J})}{9.11\times10^{-31}\ \text{kg}}} = 1.21\times10^7\ \text{m/s}.$$

SAMPLE PROBLEM 28-11. Using the expression for the electric field due to a uniform ring of positive charge at a point on its axis (the z axis), find the expression for the potential due to the ring at a point on the axis a distance z' from the ring.

Solution Equation 28-16 gives the relationship between V and $\vec{\mathbf{E}}$. The electric field for the ring of charge was given in Eq. 26-18. The field has only a z component, so the integrand of Eq. 28-16 reduces to $\vec{\mathbf{E}}\cdot d\vec{\mathbf{s}} = E_z\,dz$. We integrate in Eq. 28-16 from infinity to point P (the observation point):

$$V_P = -\int_\infty^{z'} E_z\,dz = -\int_\infty^{z'} \frac{\lambda}{2\epsilon_0}\frac{Rz}{(z^2+R^2)^{3/2}}\,dz.$$

Carrying out the integral, we find

$$V_P = \frac{1}{2\epsilon_0}\frac{\lambda R}{\sqrt{z'^2+R^2}},$$

which is identical to Eq. 28-30, obtained by integrating over the charge distribution of the ring.

28-7 CALCULATING THE FIELD FROM THE POTENTIAL

In Section 28-4 we described a method for obtaining the potential difference from the electric field. Now we discuss how to do that calculation in reverse: given the potential, we can find the electric field. That is, the double arrow connecting the two lower boxes in Table 28-1 can indeed go in either direction.

Figure 28-15a shows a positive test charge q_0 as it moves from point a (where the potential is V) to point b (potential $V + \Delta V$). In the process, the electric potential energy of q_0 changes by an amount $\Delta U = q_0\,\Delta V$. In the language of forces, we would say that there is an electric field $\vec{\mathbf{E}}$ that exerts a force $\vec{\mathbf{F}} = q_0\vec{\mathbf{E}}$ on the particle. The work done by this force as the particle moves from a to b is $W = F_s\,\Delta s = q_0 E_s\,\Delta s$, where E_s and F_s are the components of $\vec{\mathbf{E}}$ and $\vec{\mathbf{F}}$ along Δs, which represents the displacement of the particle as it travels from a to b. (We assume that Δs is small, so that we can regard the force and the field as roughly constant in both magnitude and direction along ab.) The mathematical connection between the two equivalent descriptions is $W = -\Delta U$, which gives

$$q_0 E_s\,\Delta s = -q_0\,\Delta V \qquad (28\text{-}35)$$

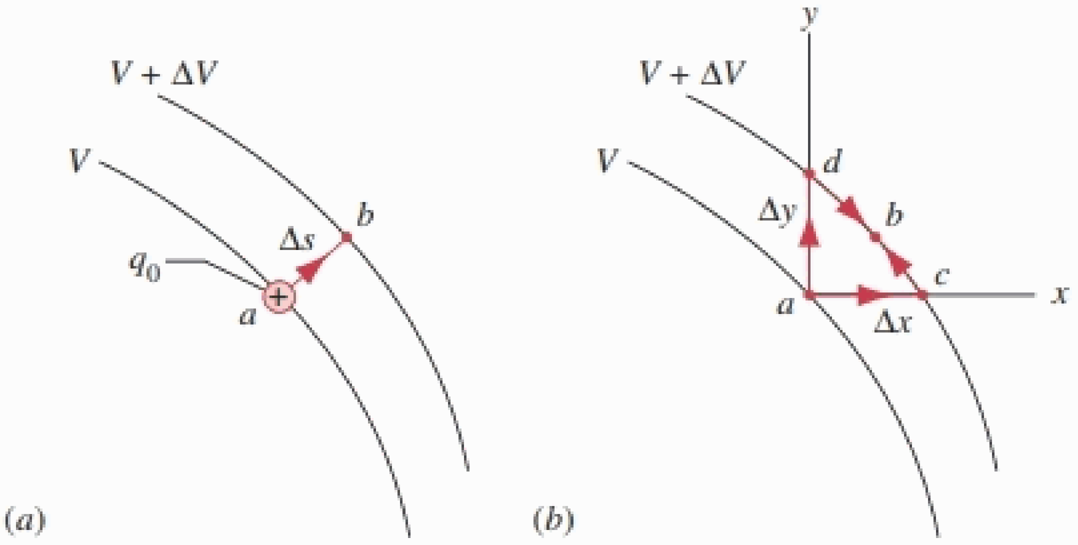


FIGURE 28-15. (a) A charged particle q_0 moves on path ab between two equipotentials. (b) The particle moves from a to b along either path acb or adb.

or

$$E_s = -\frac{\Delta V}{\Delta s}. \qquad (28\text{-}36)$$

This equation gives us the fundamental connection between the electric field and the electric potential: the electric field is the negative of the change in potential with distance. If ΔV is positive, the electric field gives a force that opposes the movement of the positively charged test particle from a to b, and if ΔV is negative, the field gives a force in the direction of the motion. In the limit of infinitesimal displacements, Eq. 28-36 becomes a derivative:

$$E_s = -\frac{dV}{ds}. \qquad (28\text{-}37)$$

The component of the electric field in any direction is the negative of the derivative of the potential with respect to a displacement in that direction.

Let us choose a different geometry for this process. Figure 28-15*b* shows the same process, but instead of moving the test charge from a to b directly, we move it along two different paths. Path *acb* takes the charge along the x axis from a to c and then along the path from c to b, which has been chosen so that the potential has the same value $V + \Delta V$ everywhere between c and b. The work done by the electric field along *cb* is zero, because the potential does not change (see Eq. 28-11). The work done by the electric field along *ac* is $F_x\,\Delta x = q_0 E_x\,\Delta x$. Because the change in potential energy is independent of path, we have again from $W = -\Delta U$

$$q_0 E_x\,\Delta x = -q_0\,\Delta V \quad \text{or} \quad E_x = -\frac{\Delta V}{\Delta x}.$$

If we move the particle on path *adb,* the work is $F_y\,\Delta y = q_0 E_y\,\Delta y$ along *ad* and zero along *db* (which has again been chosen so that the potential has the same value $V + \Delta V$ everywhere between d and b). Because the net change in potential energy along *adb* is also ΔV, we obtain

$$q_0 E_y\,\Delta y = -q_0\,\Delta V \quad \text{or} \quad E_y = -\frac{\Delta V}{\Delta y}.$$

A similar result would be obtained for E_z from a three-dimensional calculation.

If we take the limit as the path lengths become very small, the differences become derivatives, and we can write the most general relationship between $\vec{\mathbf{E}}$ and V as

$$E_x = -\frac{\partial V}{\partial x}, \quad E_y = -\frac{\partial V}{\partial y}, \quad E_z = -\frac{\partial V}{\partial z}. \qquad (28\text{-}38)$$

If $V(x, y, z)$ is known at all points in space for a particular charge distribution, then we can find the components of $\vec{\mathbf{E}}$ by taking partial derivatives of V with respect to each of the coordinates.*

We therefore have two methods of calculating the electric field for continuous charge distributions; one based on integrating Coulomb's law (Eqs. 26-13 to 26-15) and another based on differentiating the potential (Eq. 28-38). In practice, the second method often turns out to be less difficult.

SAMPLE PROBLEM 28-12. Using Eq. 28-32 for the potential on the axis of a uniformly charged disk, derive an expression for the electric field at axial points.

Solution From symmetry, $\vec{\mathbf{E}}$ must lie along the axis of the disk (the z axis). Using Eq. 28-38, we have (assuming $z > 0$)

$$\begin{aligned} E_z &= -\frac{\partial V}{\partial z} = -\frac{\sigma}{2\epsilon_0}\frac{d}{dz}[(z^2 + R^2)^{1/2} - z] \\ &= \frac{\sigma}{2\epsilon_0}\left(1 - \frac{z}{\sqrt{z^2 + R^2}}\right), \end{aligned}$$

in agreement with the result obtained by direct integration, Eq. 26-19.

SAMPLE PROBLEM 28-13. Figure 28-16 shows a (distant) point P in the field of a dipole located at the origin of an xz coordinate system. Calculate $\vec{\mathbf{E}}$ as a function of position.

Solution From symmetry, $\vec{\mathbf{E}}$ at points in the plane of Fig. 28-16 lies in this plane and can be expressed in terms of its components E_x and E_z, E_y being zero. Let us first express the potential in rectangular coordinates rather than polar coordinates, making use of

$$r = (x^2 + z^2)^{1/2} \quad \text{and} \quad \cos\theta = \frac{z}{(x^2 + z^2)^{1/2}}.$$

V is given by Eq. 28-22:

$$V = \frac{1}{4\pi\epsilon_0}\frac{p\cos\theta}{r^2}.$$

Substituting for r^2 and $\cos\theta$, we obtain

$$V = \frac{p}{4\pi\epsilon_0}\frac{z}{(x^2 + z^2)^{3/2}}.$$

We find E_z from Eq. 28-38, recalling that x is to be treated as a constant in this calculation,

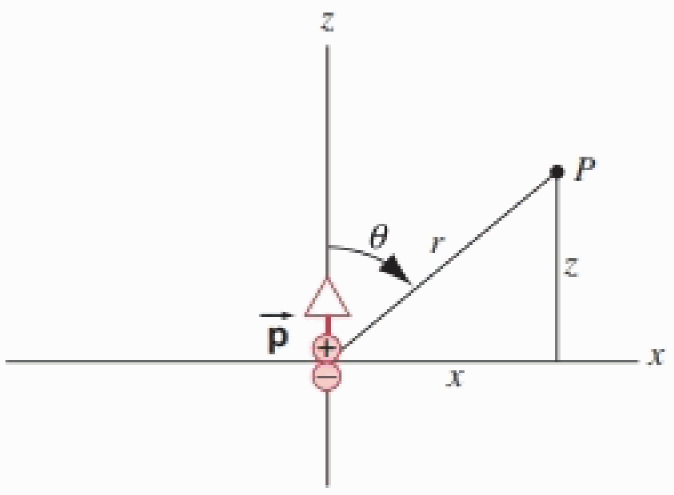


FIGURE 28-16. Sample Problem 28-13. A dipole is located at the origin of the xz system.

---

* The symbol $\partial V/\partial x$ denotes a *partial derivative.* In taking this derivative of the function $V(x, y, z)$, the quantity x is to be viewed as a variable and y and z are to be regarded as constants. Similar considerations hold for $\partial V/\partial y$ and $\partial V/\partial z$.

$$E_z = -\frac{\partial V}{\partial z} = -\frac{p}{4\pi\epsilon_0}\frac{(x^2+z^2)^{3/2} - z[\frac{3}{2}(x^2+z^2)^{1/2}](2z)}{(x^2+z^2)^3}$$

$$= -\frac{p}{4\pi\epsilon_0}\frac{x^2-2z^2}{(x^2+z^2)^{5/2}}. \qquad (28\text{-}39)$$

Putting $x = 0$ describes distant points along the dipole axis (that is, the z axis), and the expression for E_z reduces to

$$E_z = \frac{1}{4\pi\epsilon_0}\frac{2p}{z^3}.$$

This result agrees exactly with that found in Chapter 26, Problem 1 for the field along the dipole axis. Note that along the z axis, $E_x = 0$ from symmetry.

Putting $z = 0$ in Eq. 28-39 gives E_z for distant points in the median plane of the dipole:

$$E_z = -\frac{1}{4\pi\epsilon_0}\frac{p}{x^3},$$

which agrees exactly with Eq. 26-12 for, again from symmetry, E_x equals zero in the median plane. The negative sign in this equation indicates that $\vec{\mathbf{E}}$ points in the negative z direction. You can carry out a similar procedure to find E_x, and you should obtain a result that agrees with that of Problem 2 of Chapter 26.

28-8 EQUIPOTENTIAL SURFACES

Consider a point charge $q = 1.11$ nC. Using Eq. 28-18, we can find the potential due to this charge to be 100 V at a distance 0.1 m from the charge. Because there is no directionality associated with potential, its value is 100 V at that distance in any direction. This is indicated in Fig. 28-17. At any point on the sphere of radius 0.1 m surrounding q, the potential is 100 V. On a second sphere of radius 0.2 m, the potential everywhere has the value 50 V.

A surface on which the potential has the same value everywhere, such as one of the spheres in Fig. 28-17, is called an *equipotential surface.* No net work is done by electric forces when we move a test charge from any point on an equipotential surface to any other point on the same surface, because $\Delta V = 0$. Even if the path moves off the surface, no net work is done as long as the path starts and finishes on the same equipotential surface. The amount of work done by electrical forces when a test charge moves from one equipotential surface to another depends only on the potential difference between the two surfaces; the work is independent of the starting and finishing locations on the two surfaces—the same work is done when the charge moves from *any* point on the first surface to *any* point on the second surface.

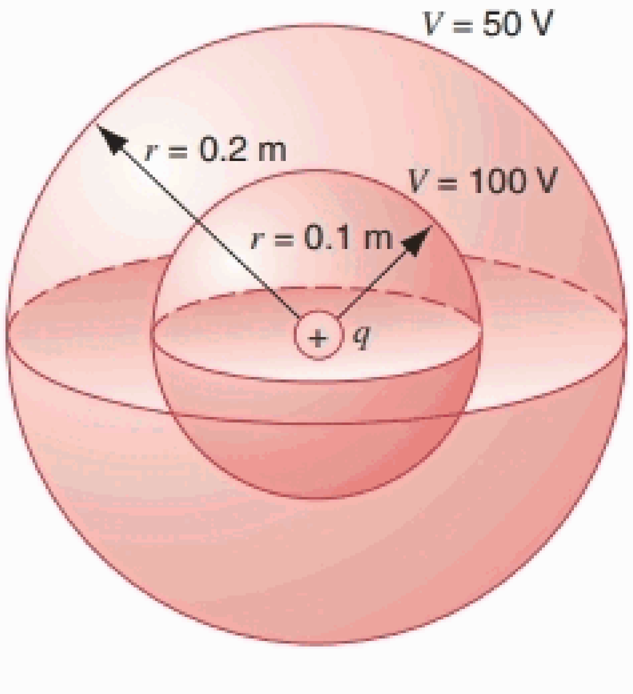


FIGURE 28-17. At all points on a sphere surrounding the charge q, the potential has the same value. Two spheres are shown, one for $V = 100$ V and another for $V = 50$ V.

Figure 28-18 shows portions of a *family* of equipotential surfaces that might be associated with a certain charge distribution. The work done by electric forces when a charged particle moves along path 1 is zero because that path starts and ends on the same equipotential surface. The work done along path 2 is zero for the same reason. The work is not zero along paths 3 and 4, but it has the same value for both these paths because they connect the same pair of equipotential surfaces. That is, paths 3 and 4 connect points with the same potential difference ($V_B - V_A$). If we move a charge q from *any* point on surface A to *any* point on surface B, the work done by the electrostatic force is, according to Eq. 28-11, $W_{AB} = -q(V_B - V_A)$.

Field Lines and Equipotential Surfaces

In Section 26-5 we discussed a different graphical method of describing a charge distribution, based on electric field lines. The mathematical relationship between $\vec{\mathbf{E}}$ and V that we derived in Section 28-7 suggests a relationship between the graphical representations as well.

Suppose we release a positive charge from rest at point b on the equipotential $V + \Delta V$ in Fig. 28-15. In the language of potential, we say that the particle would "fall" through the potential difference ΔV toward the equipotential V. We can also regard the particle as being accelerated by an electric field that is present in the region between the equipotential surfaces. The electric field must be perpendicular to the equipotential surface at point b. If this were not so, then there would be a component of the electric field along the equipotential surface, which would do work on a

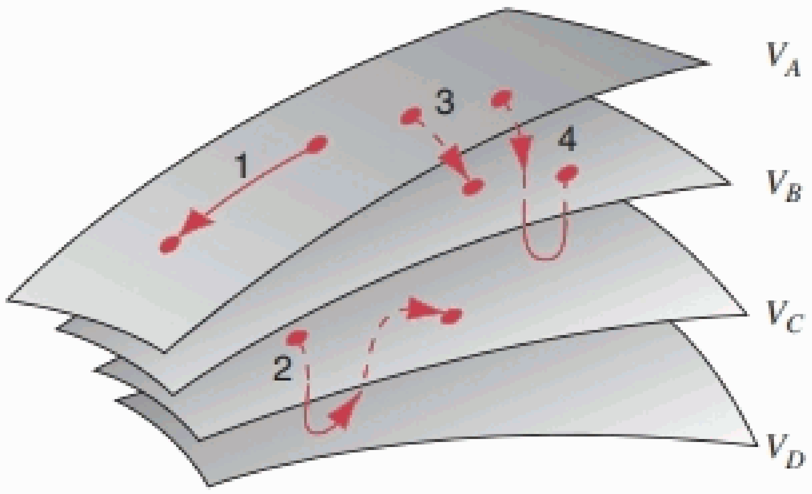


FIGURE 28-18. Portions of four equipotential surfaces. Four different paths for moving a test particle are shown.

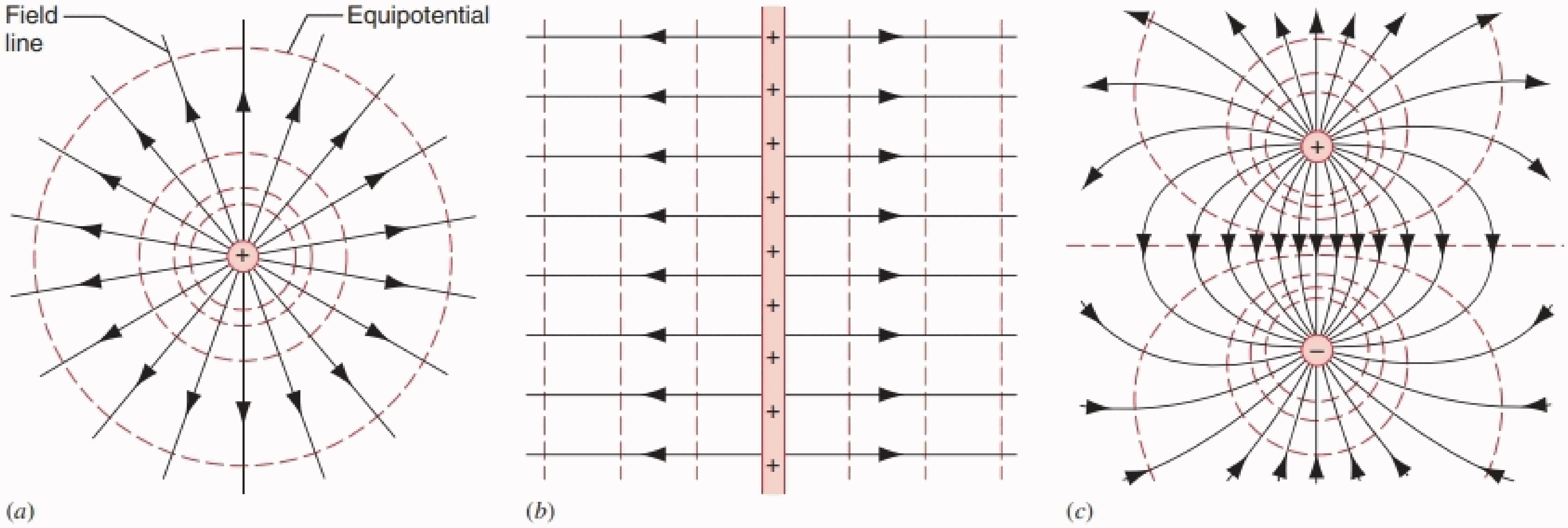


Figure 28-19. Electric field lines (solid lines) and cross sections of equipotential surfaces (dashed lines) for (*a*) a positive point charge, (*b*) an infinite sheet of positive charge, viewed along its edge, and (*c*) an electric dipole.

particle that moved along the surface. This, however, would violate the definition of an equipotential as a surface of constant potential, along which we can move a charged particle freely with no work. We conclude that *the electric field lines must everywhere be perpendicular to the equipotential surfaces.*

It is also possible to reach the same conclusion from Eq. 28-37, $E_s = -dV/ds$. There will be one direction for ds in Fig. 28-15 in which the value of the quantity $-dV/ds$ is a maximum, which means that E_s is also a maximum in that direction. That maximum value is E, the magnitude of the electric field at that point, and the direction for which E_s has its maximum is the direction of the electric field. Equivalently, we can draw at point b a circle of radius ds. One point on the circle will be the closest to the next equipotential and therefore will represent the largest value of $-dV$. The direction from b to that point is perpendicular to the equipotential surface at b and represents the direction of the electric field at b.

If we know the pattern of equipotential surfaces for a particular charge distribution, we can find the field lines by drawing perpendiculars to the equipotentials. Figure 28-19 shows the combined equipotentials and field lines for three cases we have already considered: the point charge, the infinite sheet of charge, and the dipole. These drawings represent the electric field lines of Figs. 26-10, 26-11, and 26-12 with superimposed equipotential surfaces. Note that the field lines are perpendicular to the equipotentials wherever they cross.

28-9 THE POTENTIAL OF A CHARGED CONDUCTOR

In Section 27-6 we deduced two properties of an isolated charged conductor: (1) the electric field is zero in its interior, and (2) the charge resides on the outer surface of the conductor. A third important property of a charged conductor results from considering its electric potential.

Suppose we have a conductor of arbitrary shape, to which we transfer a net charge. The charges are free to move and will quickly distribute themselves on the outer surface of the conductor until they are in equilibrium. In effect, the charges of the same sign repel one another until they reach a distribution in which the average distance between them is as large as possible, so that the potential energy of the arrangement of charges reaches a minimum value.

If the charges are in equilibrium on the surface of the conductor, then its surface must be an equipotential. If this were not so, some parts of the surface would be at higher or lower potentials than other parts. Positive charges would then migrate toward regions of low potential and negative charges toward regions of high potential. However, this contradicts our assertion that the charges are in equilibrium, and therefore the surface must be an equipotential.

If the electric field is zero in the interior of the conductor, then we may move a test charge along any path in the interior or from the surface to the interior and the net work done on the test charge by the surface charges will be zero. This means that the potential difference between any two points is zero, and thus the potential has the same value at all points in the conductor. We therefore obtain a third property of conductors: *the entire conductor is at the same potential.* This conclusion holds only in the electrostatic case; when we discuss currents flowing through conductors, a potential difference can exist between different points in the conductor.

Note that we have made no assumptions about the shape of the conductor. If the conductor is spherical, the charge is uniformly distributed over the surface. For conductors whose shape is nonspherical, the charge density is not uniform over the surface, but the surface is still an equipotential. Even in a conductor with internal cavities, whether or

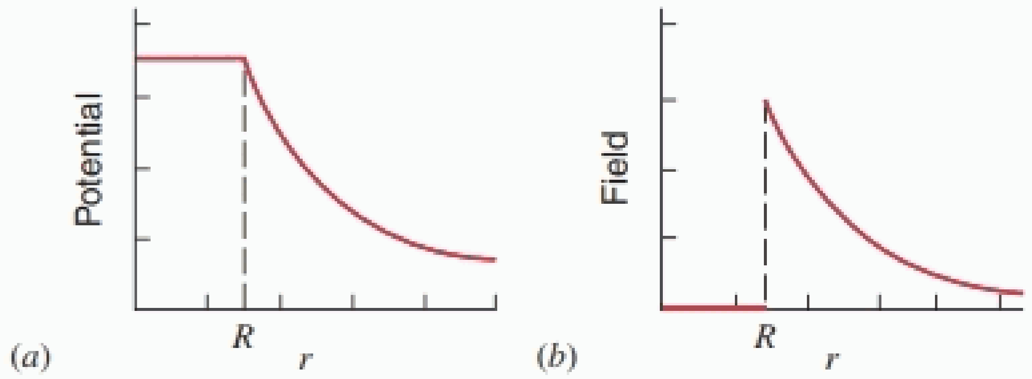


FIGURE 28-20. (*a*) The potential for a charged spherical conductor. (*b*) The electric field of the conductor.

not they contain charge, all points (surface and interior) are at the same potential.

Our conclusion about the surface of the conductor being an equipotential is consistent with the discussion in Section 28-8, in which we concluded that electric field lines are always perpendicular to equipotential surfaces. In Section 27-6 we used Gauss' law to determine that the electric field just outside the surface of the conductor is perpendicular to the surface, which must be true if the surface of the conductor is an equipotential.

We can derive explicit results for the case of a solid spherical conductor that carries a uniformly distributed total charge q on its surface. In Section 25-5, we discussed a property of a uniformly charged spherical shell: the force on an external charge is the same as if the shell were replaced with a point charge at its center. This property allows us to use the point-charge expressions for the electric potential (Eq. 28-18) and electric field (Eq. 26-6) at locations where $r > R$.

Inside the shell, the force on a point charge is zero, which means that the potential must have the same value everywhere in the conductor, including the surface. The value at the surface is found from Eq. 28-18 evaluated for $r = R$, and so the potential in the interior is

$$V = \frac{1}{4\pi\epsilon_0}\frac{q}{R}, \qquad r < R. \tag{28-40}$$

Figure 28-20 shows the field and potential for an isolated charged spherical conductor. The field is zero for $r < R$ and decreases like $1/r^2$ for $r > R$. The potential is constant for $r < R$ and falls off like $1/r$ for $r > R$.

Corona Discharge (Optional)

Although the surface charge is distributed uniformly on a spherical conductor, this will *not* be the case on conductors of arbitrary shape.* Near sharp points or edges, the surface charge density—and thus the electric field just outside the surface—can reach very high values.

To see qualitatively how this occurs, consider two conducting spheres of different radii connected by a fine wire (Fig. 28-21). Let the entire assembly be raised to some arbitrary potential V. The (equal) potentials of the two spheres, using Eq. 28-40, are

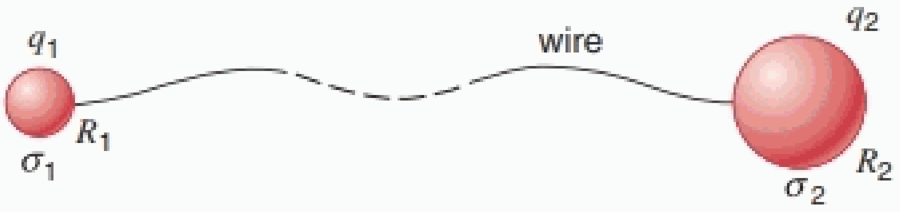


FIGURE 28-21. Two conducting spheres connected by a long fine wire.

$$V = \frac{1}{4\pi\epsilon_0}\frac{q_1}{R_1} = \frac{1}{4\pi\epsilon_0}\frac{q_2}{R_2},$$

which yields

$$\frac{q_1}{q_2} = \frac{R_1}{R_2}. \tag{28-41}$$

We assume the spheres to be so far apart that the charge on one does not affect the distribution of charge on the other.

The ratio of the surface charge densities of the two spheres is

$$\frac{\sigma_1}{\sigma_2} = \frac{q_1/4\pi R_1^2}{q_2/4\pi R_2^2} = \frac{q_1 R_2^2}{q_2 R_1^2}.$$

Combining this result with Eq. 28-41 gives

$$\frac{\sigma_1}{\sigma_2} = \frac{R_2}{R_1}. \tag{28-42}$$

Equation 28-42 suggests that the smaller sphere has the larger surface charge density. Recalling that for an external charge, the field is the same as if we replaced the sphere with a point charge at its center, we can express the field just outside the surface of the sphere as

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} = \frac{\sigma}{\epsilon_0}. \tag{28-43}$$

According to Eq. 28-42, the surface charge density is larger for the sphere of smaller radius, and thus the field is also larger just outside the sphere of smaller radius. *The smaller the radius of the sphere, the larger is the electric field just outside its surface.*

Near a sharp conductor (that is, one of very small radius) the electric field may be large enough to ionize molecules in the surrounding air; as a result the normally nonconducting air can conduct and carry charge away from the conductor. Such an effect is called a *corona discharge.* Electrostatic paint sprayers use a corona discharge to transfer charge to droplets of paint, which are then accelerated by an electric field. Photocopy machines based on the xerography process use a wire to produce a corona discharge that transfers charge to a selenium-covered surface; the charge is neutralized on regions where light strikes the surface, and the remaining charged areas attract a fine black powder that forms the image. ■

---

* See "The Lightning-rod Fallacy," by Richard H. Price and Ronald J. Crowley, *American Journal of Physics,* September 1985, p. 843, for a careful discussion of this phenomenon.

28-10 THE ELECTROSTATIC ACCELERATOR (Optional)

Many studies of nuclei involve nuclear reactions, which occur when a beam of particles is incident on a target. One method that is used to accelerate particles for nuclear reactions is based on an electrostatic technique. A particle of positive charge q "falls" through a negative change in potential ΔV and therefore experiences a negative change in its potential energy, $\Delta U = q\,\Delta V$, according to Eq. 28-14. The corresponding increase in the kinetic energy of the particle is $\Delta K = -\Delta U$, and, assuming the particle starts from rest, its final kinetic energy is

$$K = -q\,\Delta V. \tag{28-44}$$

For ionized atoms, q is normally positive. To obtain the highest energy possible for the beam, we would like to have the largest difference in potential. For applications of interest in nuclear physics, particles with kinetic energies of millions of electron-volts (MeV) are required to overcome the Coulomb force of repulsion between the incident and target particles. Kinetic energies of MeV require potential differences of millions of volts.

An electrostatic device that can produce such large potential differences is illustrated in Fig. 28-22. A small conducting sphere of radius a and carrying charge q is located inside a larger shell of radius b that carries charge Q. A conducting path is momentarily established between the two conductors, and the charge q then moves entirely to the outer conductor, no matter how much charge Q is already residing there (because the charge on a conductor always moves to its outer surface). If there is a convenient mechanism for replenishing the charge q on the inner sphere from an external supply, the charge Q on the outer sphere and its potential can, in principle, be increased without limit. In practice, the terminal potential is limited by sparking that occurs through air (Fig. 28-23).

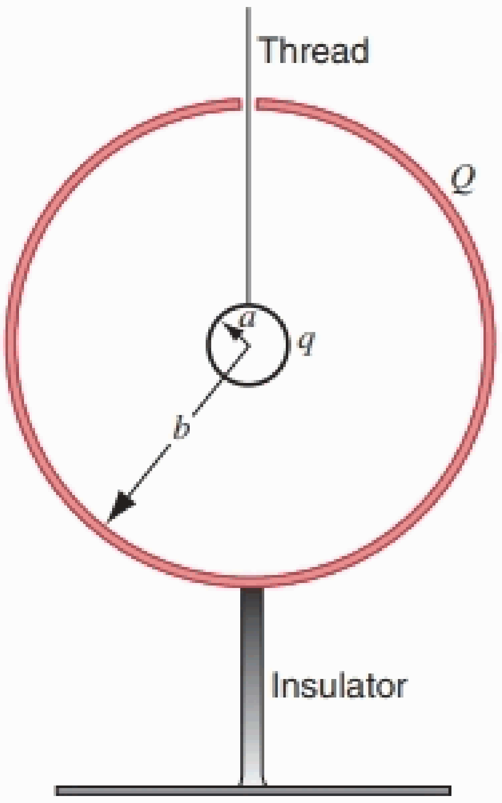


FIGURE 28-22. A small charged sphere is suspended inside a larger charged spherical shell.

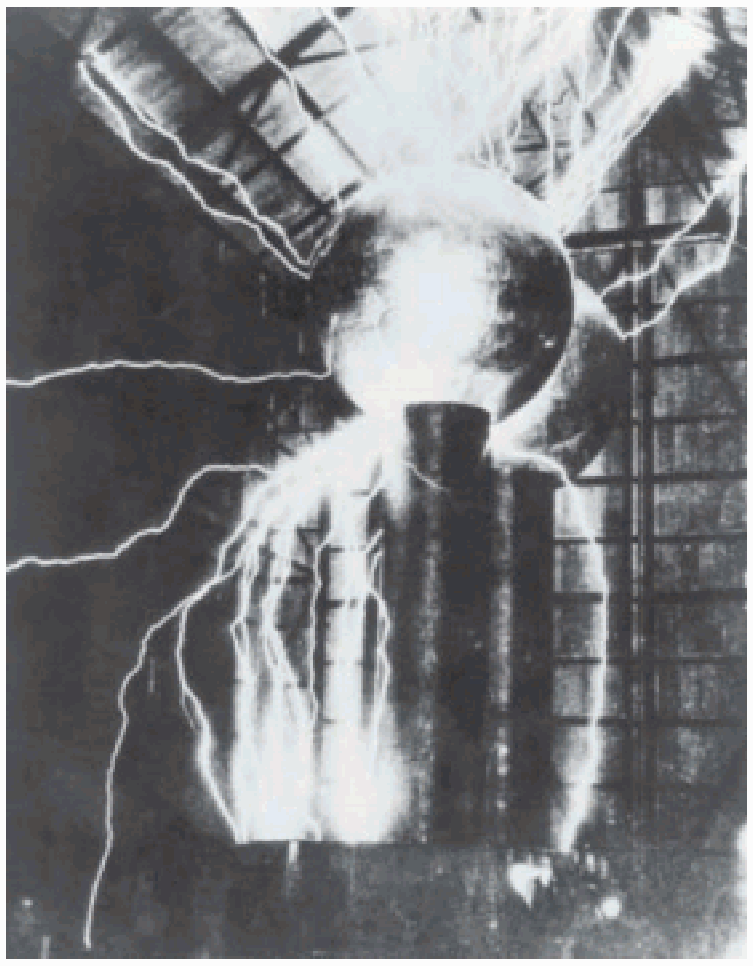


FIGURE 28-23. An electrostatic generator, with a potential of 2.7 million volts, causing sparking due to conduction through air.

This well-known principle of electrostatics was first applied to accelerating nuclear particles by Robert J. Van de Graaff in the early 1930s, and the accelerator has become known as a *Van de Graaff accelerator*. Potentials of several million volts were easily achieved, the limiting potential coming from the leakage of charge through the insulating supports or breakdown of air (or the high-pressure insulating gas) surrounding the high-voltage terminal.

Figure 28-24 shows the basic design of the Van de Graaff accelerator. Charge is sprayed from a sharp tip

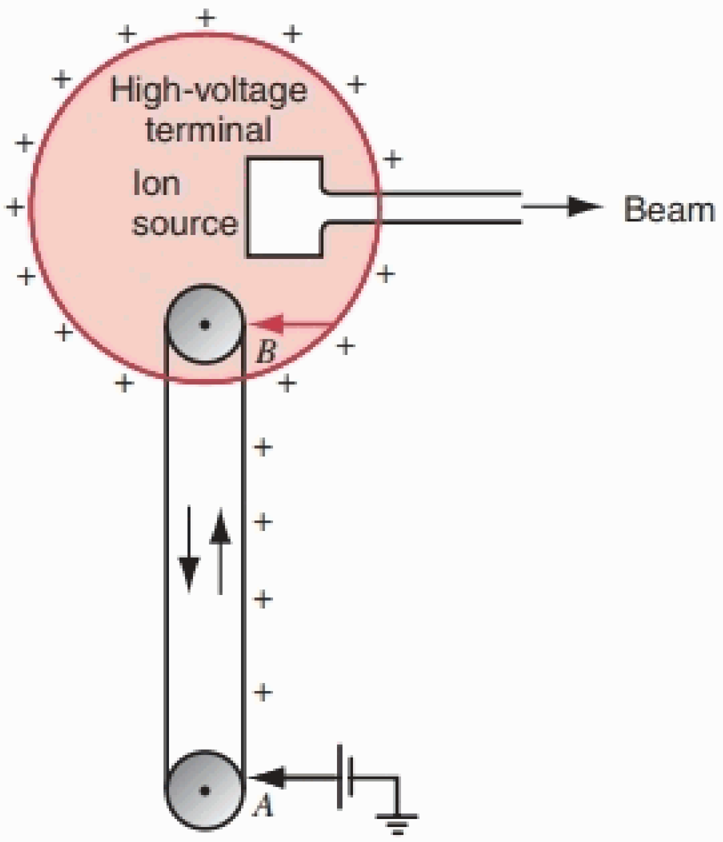


FIGURE 28-24. Diagram of Van de Graaff accelerator. Positive charge is sprayed onto the moving belt at A and is removed from the belt at B, where it flows to the terminal, which becomes charged to a potential V. Positively charged ions are repelled from the terminal to form the accelerator beam.

(called a corona point) at A onto a moving belt made of insulating material (often rubber). The belt carries the charge into the high-voltage terminal, where it is removed by another corona point B and travels to the outer conductor. Inside the terminal is a source of positive ions, for example, nuclei of hydrogen (protons) or helium (alpha particles). The ions "fall" from the high potential, gaining a kinetic energy of several MeV in the process. The terminal is enclosed in a tank that contains insulating gas to prevent sparking.

A clever variation of this basic design makes use of the same high voltage to accelerate ions twice, thereby gaining an additional increase in kinetic energy. A source of *negative* ions, made by adding an electron to a neutral atom, is located outside the terminal. These negative ions "fall toward" the positive potential of the terminal. Inside the high-voltage terminal, the beam passes through a chamber consisting of a gas or thin foil, which is designed to remove or strip several electrons from the negative ions, turning them into positive ions which then "fall from" the positive potential. Such "tandem" Van de Graaff accelerators currently use a terminal voltage of 25 million volts to accelerate ions such as carbon or oxygen to kinetic energies in excess of 100 MeV.

SAMPLE PROBLEM 28-14. Calculate the potential difference between the two spheres illustrated in Fig. 28-22.

Solution The potential difference $V(b) - V(a)$ has two contributions: one from the small sphere and one from the large spherical shell. These can be calculated independently and added algebraically. Let us first consider the large shell. Figure 28-20a shows that the potential at all interior points has the same value as the potential on the surface. Thus the contribution of the large shell to the difference $V(b) - V(a)$ is 0.

All that remains then is to evaluate the difference considering only the small sphere. For all points external to the small sphere, we can treat it as a point charge, and the potential difference can be found from Eq. 28-17:

$$V(b) - V(a) = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{b} - \frac{1}{a}\right).$$

This expression gives the difference in potential between the inner sphere and the outer shell. Note that this is *independent of the charge Q on the outer shell.* If q is positive, the difference will always be negative, indicating that the outer shell will always be at a lower potential. If positive charge is permitted to flow between the spheres, it will always flow from higher to lower potential—that is, from the inner to the outer sphere—no matter how much charge already resides on the outer spherical shell.

MULTIPLE CHOICE

28-1 Potential Energy

28-2 Electric Potential Energy

28-3 Electric Potential

1. A negative point charge is moved from a to several possible final points b in Fig. 28-25. Which path requires the greatest amount of external work to move the particle?

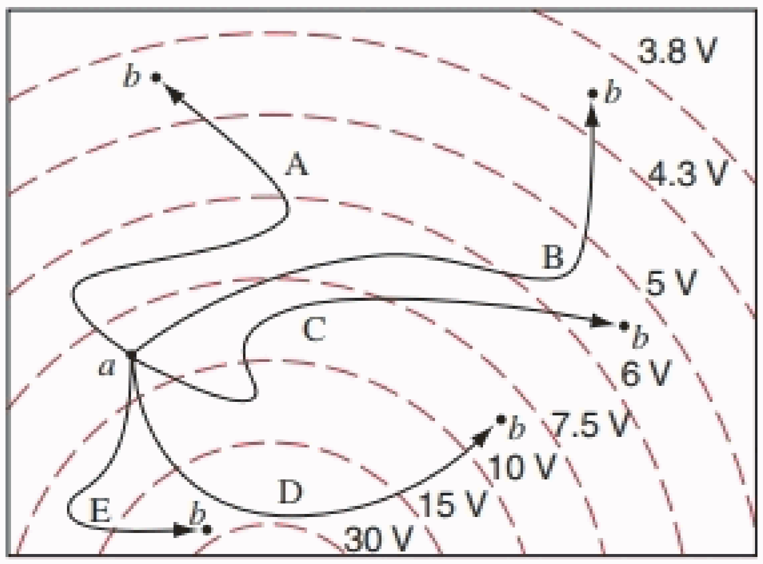


FIGURE 28-25. Multiple-choice question 1.

2. An electron is released from rest in a region of space with a nonzero electric field. Which of the following statements is true?
   (A) The electron will begin moving toward a region of higher potential.
   (B) The electron will begin moving toward a region of lower potential.
   (C) The electron will begin moving along a line of constant potential.
   (D) Nothing can be concluded unless the direction of the electric field in known.

28-4 Calculating the Potential from the Field

3. Inside a charged conductor under electrostatic conditions,
   (A) $V = 0$. (B) $\partial V/\partial x = 0$. (C) $\partial^2 V/\partial x^2 = 0$.
   (D) Two of (A), (B), or (C) must be true.
   (E) All three must be true.
4. The electric field lines are closer together near object A than they are near object B. We can conclude
   (A) the potential near A is greater than the potential near B.
   (B) the potential near A is less than the potential near B.
   (C) the potential near A is equal to the potential near B.
   (D) nothing about the relative potentials near A or B.
5. Figure 28-26 shows the electric field lines around three point charges, A, B, and C.
   (*a*) Which point corresponds to the highest potential?
   (A) P (B) Q (C) R
   (D) All three points are at the same potential.
   (*b*) Which point corresponds to the lowest potential?
   (A) P (B) Q (C) R
   (D) All three points are at the same potential.

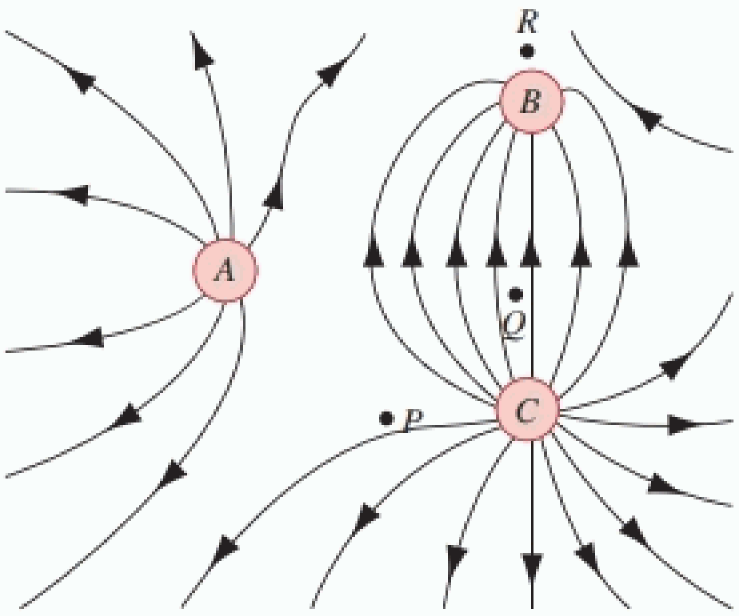


FIGURE 28-26. Multiple-choice question 5.

28-5 Potential Due to Point Charges

6. A single positive point charge q is located as shown in Fig. 28-27*a*, and the potential at point P is V_0 (with $V = 0$ at infinity).

(*a*) A second charge $q' = +q$ is placed equidistant from P as shown in Fig. 28-27*b*. The potential at P is now

(A) $4V_0$. (B) $2V_0$. (C) $\sqrt{2}V_0$.
(D) $V_0/2$. (E) 0.

(*b*) Instead of a positive charge, a negative charge $q' = -q$ is located as shown in Fig. 28-27*b*. The potential at P is now

(A) $4V_0$. (B) $2V_0$. (C) $\sqrt{2}V_0$.
(D) $V_0/2$. (E) 0.

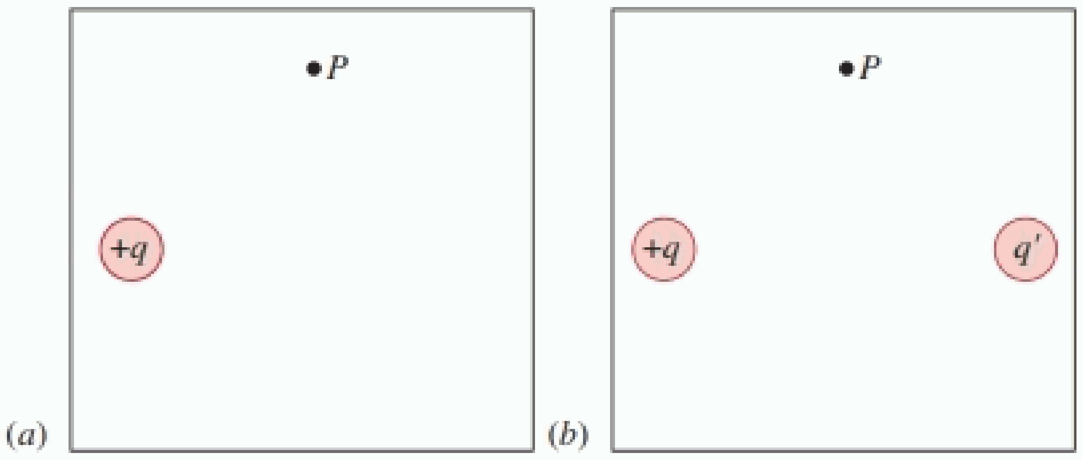


FIGURE 28-27. Multiple-choice question 6.

7. It requires 1 mJ of work to move two identical positive charges $+q$ from infinity so that they are separated by a distance a.

(*a*) How much work is required to move three identical positive charges $+q$ from infinity so that they are arranged at the vertices of an equilateral triangle with edge length a?

(A) 2 mJ (B) 3 mJ (C) 4 mJ (D) 9 mJ

(*b*) How much work is required to move four identical positive charges $+q$ from infinity so that they are arranged at the vertices of a tetrahedron with edge length a?

(A) 3 mJ (B) 4 mJ (C) 6 mJ (D) 16 mJ

8. A point charge $+q$ is located at the origin, and a point charge $+2q$ is located at $x = a$, where a is positive; here $V(\infty) = 0$.

(*a*) Which of the following statements is true?

(A) Close to the charges, the electric potential can be zero off the x axis.
(B) The magnitude of the electric potential will be a maximum on the x axis.
(C) The electric potential can be zero in the region between the charges.
(D) The electric potential can be zero only on the x axis.

(*b*) In which of the following regions on the x axis might there exist a point where the electric potential is zero?

(A) $-\infty < x < 0$ (B) $0 < x < a$
(C) $a < x < \infty$
(D) V does not vanish in the region $-\infty < x < \infty$.

9. A point charge $+q$ is located at the origin, and a point charge $-2q$ is located at $x = a$, where a is positive; here $V(\infty) = 0$.

(*a*) Which of the following statements is true?

(A) Close to the charges, the electric potential can be zero off the x axis.
(B) Close to the charges, the magnitude of the electric potential can be a maximum off the x axis.
(C) The electric potential can be zero only between the charges.
(D) The electric potential can be zero only on the x axis.

(*b*) In which of the following region or regions might there exist a point where the electric potential is zero?

(A) $-\infty < x < 0$ (B) $0 < x < a$
(C) $a < x < \infty$
(D) V does not vanish in the region $-\infty < x < \infty$.

28-6 Electric Potential of Continuous Charge Distributions

10. Consider the electric potential $V(z)$ on the axis of a uniform ring of positive charge; here $V(\infty) = 0$.

(*a*) $V(z)$ will have its largest value where

(A) $z = 0$. (B) $0 < |z| < \infty$.
(C) $|z| = \infty$. (D) (A) and (C) are correct.

(*b*) $|V(z)|$ can be zero where

(A) $z = 0$. (B) $0 < |z| < \infty$.
(C) $|z| = \infty$. (D) (A) and (C) are correct.

11. Consider the electric potential $V(z)$ on the axis of a uniform disk of positive charge; here $V(\infty) = 0$.

(*a*) $V(z)$ will have its largest value where

(A) $z = 0$. (B) $0 < |z| < \infty$.
(C) $|z| = \infty$. (D) (A) and (C) are correct.

(*b*) $V(z)$ can be zero where

(A) $z = 0$. (B) $0 < |z| < \infty$.
(C) $|z| = \infty$. (D) (A) and (C) are correct.

28-7 Calculating the Field from the Potential

12. A small positive charge located at the origin experiences an electrostatic force directed along the x axis. One can conclude that at the origin

(A) $V \neq 0$. (B) $\partial V/\partial x \neq 0$. (C) $\partial^2 V/\partial x^2 \neq 0$.
(D) Two of (A), (B), or (C) must be true.
(E) All three must be true.

13. An electric dipole parallel to the x axis and located at the origin experiences an electrostatic force directed along the x axis. One can conclude that at the origin

(A) $V \neq 0$. (B) $\partial V/\partial x \neq 0$. (C) $\partial^2 V/\partial x^2 \neq 0$.
(D) Two of (A), (B), or (C) must be true.
(E) All three must be true.

28-8 Equipotential Surfaces

14. Which of the following is always true for the electric flux Φ_E through a closed equipotential surface?

(A) $\Phi_E = 0$ (B) $\Phi_E > 0$ (C) $\Phi_E < 0$

(D) Φ_E is proportional to the net charge inside the surface.

28-9 The Potential of a Charged Conductor

15. A small conducting sphere originally has a charge $+q$. The sphere is lowered into a conducting can. (*a*) Which of the following quantities are fixed as the sphere is lowered, but before it touches the can? (There may be more than one correct answer.)

(A) The potential of the can
(B) The potential of the sphere
(C) The charge on the sphere
(D) The net charge on the sphere and can

(*b*) The ball touches the can. Which of the following quantities are the same both before and after the ball touches the can? (There may be more than one correct answer.)

(A) The potential of the can
(B) The potential of the sphere
(C) The charge on the sphere
(D) The net charge on the sphere and can

16. Two small conducting spheres ($r = 1$ cm) separated by a distance of 1 m each have identical positive charges. The electric potential of one sphere is V_0 (with $V = 0$ at infinity). (*a*) The potential of the other sphere

(A) is greater than V_0.
(B) is less than V_0.
(C) is equal to V_0.
(D) cannot be determined without additional information.

(*b*) The spheres are brought closer together until they touch. The electric potential of the two spheres is now V, where

(A) $V = V_0$. (B) $V_0 < V < 2V_0$.
(C) $V = 2V_0$. (D) $2V_0 < V$.

28-10 The Electrostatic Accelerator

QUESTIONS

1. Are we free to call the potential of the Earth $+100$ V instead of zero? What effect would such an assumption have on measured values of (*a*) potentials and (*b*) potential differences?

2. What would happen to you if you were on an insulated stand and your potential were increased by 10 kV with respect to the Earth?

3. Why is the electron-volt often a more convenient unit of energy than the joule?

4. How would a proton-volt compare with an electron-volt? The mass of a proton is 1840 times that of an electron.

5. Does the amount of work per unit charge required to transfer electric charge from one point to another in an electrostatic field depend on the amount of charge transferred?

6. Distinguish between potential difference and difference of potential energy. Give examples of statements in which each term is used properly.

7. Estimate the combined energy of all the electrons striking the screen of a cathode ray oscilloscope in 1 second.

8. Why is it possible to shield a room against electrical forces but not against gravitational forces?

9. Suppose the Earth has a net charge that is not zero. Why is it still possible to adopt the Earth as a standard reference point of potential and to assign the potential $V = 0$ to it?

10. Can there be a potential difference between two conductors that carry like charges of the same magnitude?

11. Give examples of situations in which the potential of a charged body has a sign opposite to that of its charge.

12. Can two different equipotential surfaces intersect?

13. An electrical worker was accidentally electrocuted and a newspaper account reported: "He accidentally touched a high-voltage cable and 20,000 V of electricity surged through his body." Criticize this statement.

14. Advice to mountaineers caught in lightning and thunderstorms is (*a*) get rapidly off peaks and ridges, and (*b*) put both feet together and crouch in the open, only the feet touching the ground. What is the basis of this good advice?

15. If $\vec{E}$ equals zero at a given point, must V equal zero for that point? Give some examples to prove your answer.

16. If you know $\vec{E}$ only at a given point, can you calculate V at that point? If not, what additional information do you need?

17. In Fig. 28-18, is the electric field E greater at the left or at the right of the figure?

18. Is the uniformly charged, nonconducting disk of Sample Problem 28-12 a surface of constant potential? Explain.

19. We have seen that, inside a hollow conductor, you are shielded from the fields of outside charges. If you are outside a hollow conductor that contains charges, are you shielded from the fields of these charges? Explain why or why not.

20. If the surface of a charged conductor is an equipotential, does that mean that charge is distributed uniformly over that surface? If the electric field is constant in magnitude over the surface of a charged conductor, does that mean that the charge is distributed uniformly?

21. In Section 28-9 we were reminded that charge delivered to the inside of an isolated conductor is transferred entirely to the outer surface of the conductor, no matter how much charge is already there. Can you keep this up forever? If not, what stops you?

22. Why can an isolated atom not have a permanent electric dipole moment?

23. Ions and electrons act like condensation centers; water droplets form around them in air. Explain why.

24. If V equals a constant throughout a given region of space, what can you say about $\vec{E}$ in that region?

25. In Chapter 14 we saw that the gravitational field strength is zero inside a spherical shell of matter. The electrical field strength is zero not only inside an isolated charged spherical conductor but inside an isolated conductor of any shape. Is the gravitational field strength inside, say, a cubical shell of matter zero? If not, in what respect is the analogy not complete?

26. How can you ensure that the electric potential in a given region of space will have a constant value?

27. Devise an arrangement of three point charges, separated by finite distances, that has zero electric potential energy.

28. A charge is placed on an insulated conductor in the form of a perfect cube. What will be the relative charge density at various points on the cube (surfaces, edges, corners)? What will happen to the charge if the cube is in air?

29. We have seen (Section 28-9) that the potential inside a conductor is the same as that on its surface. (*a*) What if the conductor is irregularly shaped and has an irregularly shaped cavity inside? (*b*) What if the cavity has a small "worm hole" connecting it to the outside? (*c*) What if the cavity is closed but has a point charge suspended within it? Discuss the potential within the conducting material and at different points within the cavities.

30. An isolated conducting spherical shell carries a negative charge. What will happen if a positively charged metal object is placed in contact with the shell interior? Discuss the three cases in which the positive charge is (*a*) less than, (*b*) equal to, and (*c*) greater than the negative charge in magnitude.

31. An uncharged metal sphere suspended by a silk thread is placed in a uniform external electric field. What is the magnitude of the electric field for points inside the sphere? Is your answer changed if the sphere carries a charge?

EXERCISES

28-1 Potential Energy

28-2 Electric Potential Energy

1. In the quark model of fundamental particles, a proton is composed of three quarks: two "up" quarks, each having charge $+\frac{2}{3}e$, and one "down" quark, having charge $-\frac{1}{3}e$. Suppose that the three quarks are equidistant from each other. Take the distance to be 1.32×10^{-15} m and calculate (*a*) the potential energy of the interaction between the two "up" quarks and (*b*) the total electric potential energy of the system.

2. Derive an expression for the work required by an external agent to put the four charges together as indicated in Fig. 28-28. Each side of the square has length a.

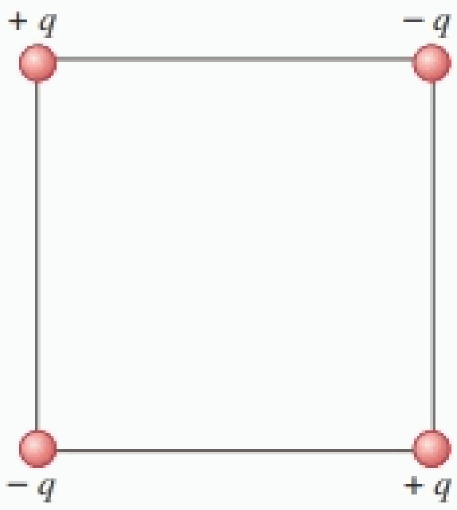


FIGURE 28-28. Exercise 2.

3. A decade before Einstein published his theory of relativity, J. J. Thomson proposed that the electron might be made up of small parts and that its mass is due to the electrical interaction of the parts. Furthermore, he suggested that the energy equals mc^2. Make a rough estimate of the electron mass in the following way: assume that the electron is composed of three identical parts that are brought in from infinity and placed at the vertices of an equilateral triangle having sides equal to the classical radius of the electron, 2.82×10^{-15} m. (*a*) Find the total electric potential energy of this arrangement. (*b*) Divide by c^2 and compare your result to the accepted electron mass (9.11×10^{-31} kg). The result improves if more parts are assumed; see Problem 2. Today, the electron is thought to be a single, indivisible particle.

4. The charges shown in Fig. 28-29 are fixed in space. Find the value of the distance x so that the electric potential energy of the system is zero.

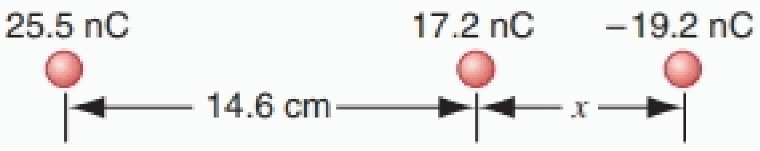


FIGURE 28-29. Exercise 4.

5. Figure 28-30 shows an idealized representation of a ^{238}U nucleus ($Z = 92$) on the verge of fission. Calculate (*a*) the repulsive force acting on each fragment and (*b*) the mutual electric potential energy of the two fragments. Assume that the fragments are equal in size and charge, spherical, and just touching. The radius of the initially spherical ^{238}U nucleus is 8.0 fm. Assume that the material out of which nuclei are made has a constant density.

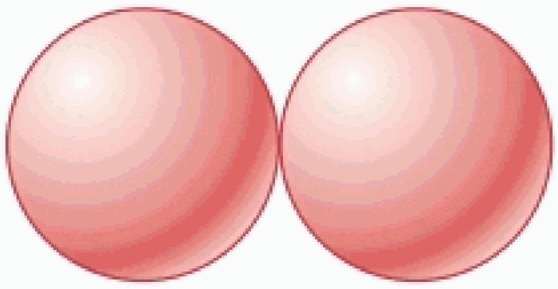


FIGURE 28-30. Exercise 5.

28-3 Electric Potential

6. Two parallel, flat, conducting surfaces of spacing $d = 1.0$ cm have a potential difference ΔV of 10.3 kV. An electron is projected from one plate directly toward the second. What is the initial velocity of the electron if it comes to rest just at the surface of the second plate? Ignore relativistic effects.

7. In a typical lightning flash, the potential difference between discharge points is about 1.0×10^9 V and the quantity of charge transferred is about 30 C. (*a*) How much energy is released? (*b*) If all the energy released could be used to accelerate a 1200-kg automobile from rest, what would be the final speed of the automobile? (*c*) If it could be used to melt ice, how much ice would it melt at 0°C?

8. The electric potential difference between discharge points during a particular thunderstorm is 1.23×10^9 V. What is the magnitude of the change in the electric potential energy of an electron that moves between these points? Give your answer in (*a*) joules and (*b*) electron-volts.

9. A particle of charge q is kept in a fixed position at a point P and a second particle of mass m, having the same charge q, is initially held at rest a distance r_1 from P. The second particle is then released and is repelled from the first one. Determine its speed at the instant it is a distance r_2 from P. Let $q = 3.1\ \mu\text{C}$, $m = 18$ mg, $r_1 = 0.90$ mm, and $r_2 = 2.5$ mm.

10. An electron is projected with an initial speed of 3.44×10^5 m/s directly toward a proton that is essentially at rest. If the electron is initially a great distance from the proton, at what distance from the proton is its speed instantaneously equal to twice its initial value?

11. Calculate (*a*) the electric potential established by the nucleus of a hydrogen atom at the average distance of the circulating electron ($r = 5.29 \times 10^{-11}$ m); (*b*) the electric potential energy of the atom when the electron is at this radius; and (*c*) the kinetic energy of the electron, assuming it to be moving in a circular orbit of this radius centered on the nucleus. (*d*) How much energy is required to ionize the hydrogen atom? Express all energies in electron-volts, and take $V = 0$ at infinity.

12. In the rectangle shown in Fig. 28-31, the sides have lengths 5.0 cm and 15 cm, $q_1 = -5.0\ \mu\text{C}$, and $q_2 = +2.0\ \mu\text{C}$. (*a*) What are the electric potentials at corner B and at corner A? (Assume $V = 0$ at infinity.) (*b*) How much external work is required to move a third charge $q_3 = +3.0\ \mu\text{C}$ from B to A along a diagonal of the rectangle? (*c*) In this process, is the external work converted into electrostatic potential energy or vice versa? Explain.

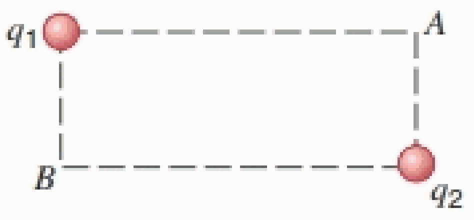


FIGURE 28-31. Exercise 12.

28-4 Calculating the Potential from the Field

13. Two large, parallel, conducting plates are 12.0 cm apart and carry equal but opposite charges on their facing surfaces. An electron placed midway between the two plates experiences a force of 3.90×10^{-15} N. (*a*) Find the electric field at the position of the electron. (*b*) What is the potential difference between the plates?

14. An infinite sheet of charge has a charge density $\sigma = 0.12\ \mu\text{C/m}^2$. How far apart are the equipotential surfaces whose potentials differ by 48 V?

15. A Geiger counter has a metal cylinder 2.10 cm in diameter along whose axis is stretched a wire 1.34×10^{-4} cm in diameter. If 855 V is applied between them, find the electric field at the surface of (*a*) the wire and (*b*) the cylinder. (Hint: Use the result of Problem 10, Chapter 27.)

16. In the Millikan oil-drop experiment (see Section 26-6), an electric field of 1.92×10^5 N/C is maintained at balance across two plates separated by 1.50 cm. Find the potential difference between the plates.

28-5 Potential Due to Point Charges

17. A gold nucleus contains a positive charge equal to that of 79 protons and has a radius of 7.0 fm; see Sample Problem 28-7. An alpha particle (which consists of two protons and two neutrons) has a kinetic energy K at points far from the nucleus and is traveling directly toward it. The alpha particle just touches the surface of the nucleus where its velocity is reversed in direction. (*a*) Calculate K. (*b*) The actual alpha particle energy used in the experiment of Rutherford and his collaborators that led to the discovery of the concept of the atomic nucleus was 5.0 MeV. What do you conclude?

18. Compute the escape speed for an electron from the surface of a uniformly charged sphere of radius 1.22 cm and total charge 1.76×10^{-15} C. Neglect gravitational forces.

19. A point charge has $q = +1.16\ \mu\text{C}$. Consider point A, which is 2.06 m distant, and point B, which is 1.17 m distant in a direction diametrically opposite, as in Fig. 28-32*a*. (*a*) Find the potential difference $V_A - V_B$. (*b*) Repeat if points A and B are located as in Fig. 28-32*b*.

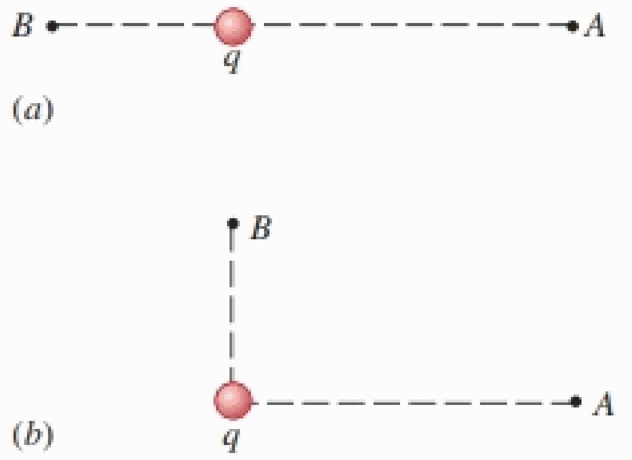


FIGURE 28-32. Exercise 19.

20. Much of the material comprising Saturn's rings (see Fig. 28-33) is in the form of tiny dust grains having radii on the order of 1.0 μm. These grains are in a region containing a dilute ionized gas, and they pick up excess electrons. If the electric potential at the surface of a grain is -400 V (relative to $V = 0$ at infinity), how many excess electrons has it picked up?

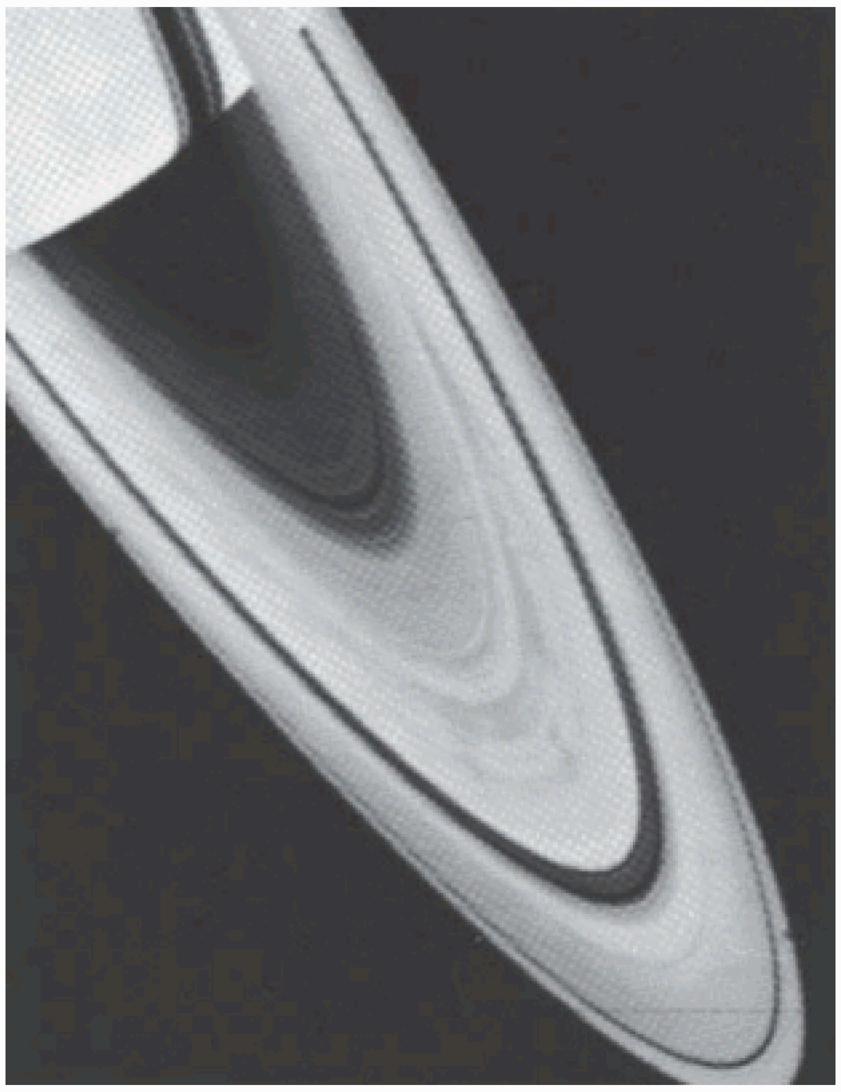


FIGURE 28-33. Exercise 20.

21. As a space shuttle moves through the dilute ionized gas of the Earth's ionosphere, its potential is typically changed by -1.0 V before it completes one revolution. By assuming that the shuttle is a sphere of radius 10 m, estimate the amount of charge it collects.

22. Suppose that the negative charge in a copper one-cent coin were removed to a very large distance from the Earth—perhaps to a distant galaxy—and that the positive charge were distributed uniformly over the Earth's surface. By how much would the electric potential at the surface of the Earth change? (See Sample Problem 25-1.)

23. An electric field of approximately 100 V/m is often observed near the surface of the Earth. If this field were the same over the entire surface, what would be the electric potential of a point on the surface? Assume that $V = 0$ at infinity.

24. The ammonia molecule NH_3 has a permanent electric dipole moment equal to 1.47 D, where D is the debye unit with a value of 3.34×10^{-30} C·m. Calculate the electric potential due to an ammonia molecule at a point 52.0 nm away along the axis of the dipole. Assume $V = 0$ at infinity.

25. (*a*) For Fig. 28-34, derive an expression for $V_A - V_B$. (*b*) Does your result reduce to the expected answer when $d = 0$? When $a = 0$? When $q = 0$?

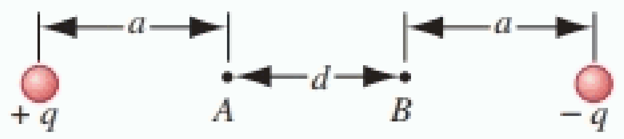


FIGURE 28-34. Exercise 25.

26. In Fig. 28-35, locate the points, if any, (*a*) where $V = 0$ and (*b*) where $E = 0$. Consider only points on the axis and take $V = 0$ at infinity.

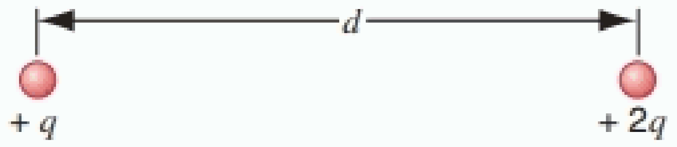


FIGURE 28-35. Exercise 26.

27. Two charges $q = +2.13\ \mu\text{C}$ are fixed in space a distance $d = 1.96$ cm apart, as shown in Fig. 28-36. (*a*) What is the electric potential at point C? Take $V = 0$ at infinity. (*b*) You bring a third charge $Q = +1.91\ \mu\text{C}$ slowly from infinity to C. How much work must you do? (*c*) What is the potential energy U of the configuration when the third charge is in place?

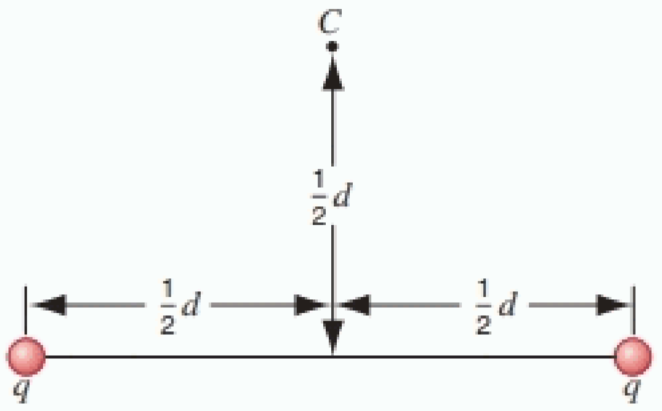


FIGURE 28-36. Exercise 27.

28-6 Electric Potential of Continuous Charge Distributions

28. At what distance along the axis of a uniformly charged disk of radius R is the electric potential equal to one-half the value of the potential at the surface of the disk at the center?

29. An electric charge of -9.12 nC is uniformly distributed around a ring of radius 1.48 m that lies in the yz plane with its center at the origin. A particle carrying a charge of -5.93 pC is located on the x axis at $x = 3.07$ m. Calculate the work done by an external agent in moving the point charge to the origin.

28-7 Calculating the Field from the Potential

30. Suppose that the electric potential varies along the x axis as shown in the graph of Fig. 28-37. Of the intervals shown (ignore the behavior at the endpoints of the intervals), determine the intervals in which E_x has (*a*) its greatest absolute value and (*b*) its least. (*c*) Plot E_x versus x.

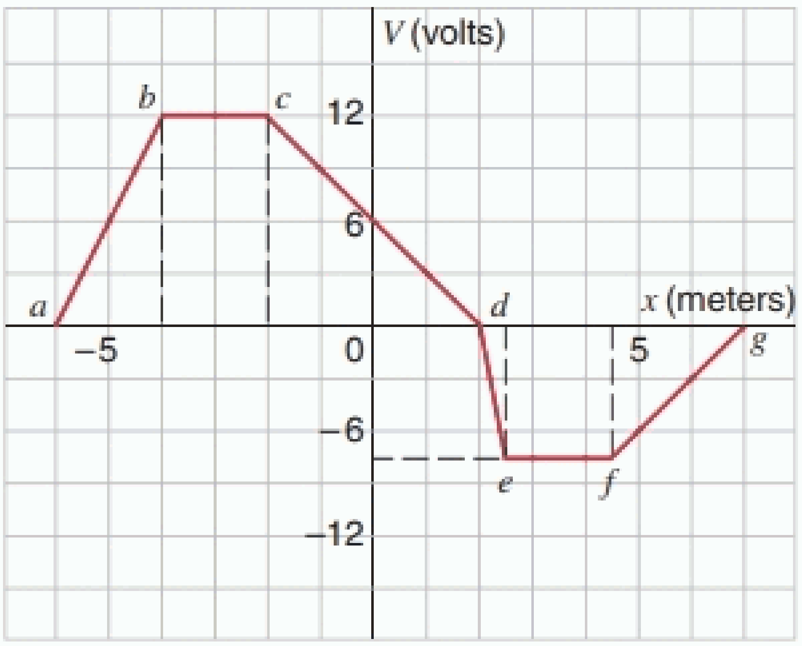


FIGURE 28-37. Exercise 30.

31. Two large, parallel, metal plates are 1.48 cm apart and carry equal but opposite charges on their facing surfaces. The negative plate is grounded and its potential is taken to be zero. If the potential halfway between the plates is $+5.52$ V, what is the electric field in this region?

32. From Eq. 28-30 derive an expression for E at axial points of a uniformly charged ring.

33. Calculate the radial potential gradient, $\partial V/\partial r$, at the surface of a gold nucleus. See Sample Problem 28-7.

34. Exercise 39 of Chapter 26 deals with Rutherford's calculation of the electric field a distance r from the center of an atom. He also gave the electric potential as

$$V = \frac{Ze}{4\pi\epsilon_0}\left(\frac{1}{r} - \frac{3}{2R} + \frac{r^2}{2R^3}\right).$$

(*a*) Show how the expression for the electric field given in Exercise 39 of Chapter 26 follows from this expression for V. (*b*) Why does this expression for V not go to zero as $r \to \infty$?

35. The electric potential V in the space between the plates of a particular, and now obsolete, vacuum tube is given by $V = (1530\ \text{V/m}^2)x^2$, where x is the distance from one of the plates. Calculate the magnitude and direction of the electric field at $x = 1.28$ cm.

28-8 Equipotential Surfaces

36. Two line charges are parallel to the z axis. One, of charge per unit length $+\lambda$, is a distance a to the right of this axis. The other, of charge per unit length $-\lambda$, is a distance a to the left of this axis (the lines and the z axis being in the same plane). Sketch some of the equipotential surfaces.

37. In moving from A to B along an electric field line, the electric field does 3.94×10^{-19} J of work on an electron in the field illustrated in Fig. 28-38. What are the differences in the electric potential (*a*) $V_B - V_A$, (*b*) $V_C - V_A$, and (*c*) $V_C - V_B$?

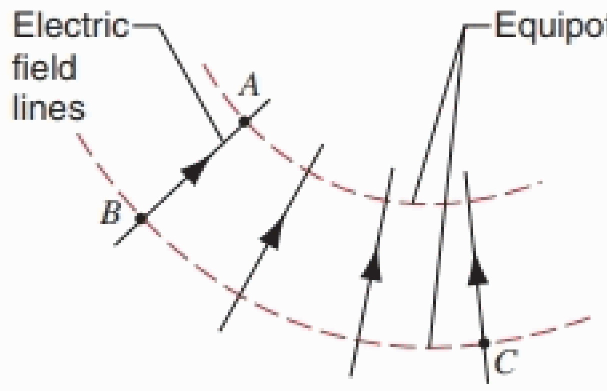


FIGURE 28-38. Exercise 37.

38. Consider a point charge with $q = 1.5 \times 10^{-8}$ C. (*a*) What is the radius of an equipotential surface having a potential of 30 V? Take $V = 0$ at infinity. (*b*) Are surfaces whose potentials differ by a constant amount (1.0 V, say) evenly spaced?

39. In Fig. 28-39 sketch qualitatively (*a*) the electric field lines and (*b*) the intersections of the equipotential surfaces with the plane of the figure. (Hint: Consider the behavior close to each point charge and at considerable distances from the pair of charges.)

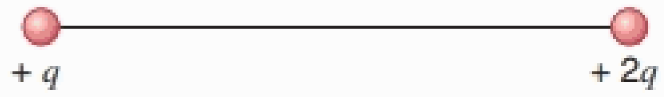


FIGURE 28-39. Exercise 39.

40. Three long parallel lines of charge have the relative linear charge densities shown in Fig. 28-40. Sketch some electric field lines and the intersections of some equipotential surfaces with the plane of this figure.

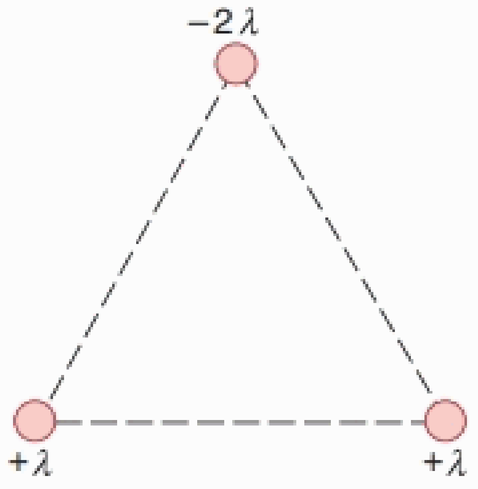


FIGURE 28-40. Exercise 40.

28-9 The Potential of a Charged Conductor

41. A thin, conducting, spherical shell of outer radius 20 cm carries a charge of $+3.0\ \mu$C. Sketch (*a*) the magnitude of the electric field $\vec{\mathbf{E}}$ and (*b*) the potential V versus the distance r from the center of the shell. (Set $V = 0$ at infinity.)

42. Consider two widely separated conducting spheres, 1 and 2, the second having twice the diameter of the first. The smaller sphere initially has a positive charge q and the larger one is initially uncharged. You now connect the spheres with a long thin wire. (*a*) How are the final potentials V_1 and V_2 of the spheres related? (*b*) Find the final charges q_1 and q_2 on the spheres in terms of q.

43. (*a*) If the Earth had a net charge equivalent to 1 electron/m^2 of surface area (a very artificial assumption), what would be the Earth's potential? (Set $V = 0$ at infinity.) (*b*) What would be the electric field due to the Earth just outside its surface?

44. A charge of 15 nC can be produced by simple rubbing. To what potential (relative to $V = 0$ at infinity) would such a charge raise an isolated conducting sphere of 16 cm radius?

45. Find (*a*) the charge and (*b*) the charge density on the surface of a conducting sphere of radius 15.2 cm whose potential is 215 V. Take $V = 0$ at infinity.

46. The metal object in Fig. 28-41 is a figure of revolution about the horizontal axis. If it is charged negatively, sketch roughly a few equipotentials and electric field lines. Use physical reasoning rather than mathematical analysis.

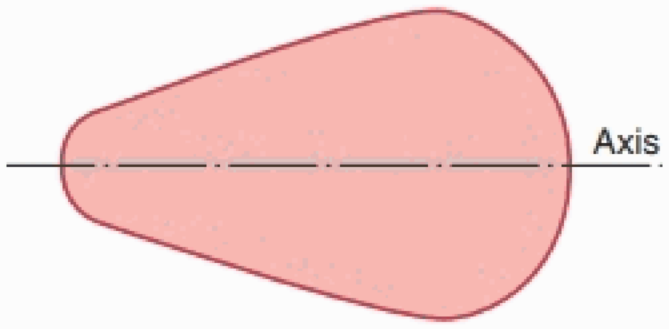


FIGURE 28-41. Exercise 46.

47. Two conducting spheres, one of radius 5.88 cm and the other of radius 12.2 cm, each have a charge of 28.6 nC and are very far apart. If the spheres are subsequently connected by a conducting wire, find (*a*) the final charge on and (*b*) the potential of each sphere, assuming $V = 0$ at infinity.

48. A charged metal sphere of radius 16.2 cm has a net charge of 31.5 nC. (*a*) Find the electric potential at the sphere's surface, if $V = 0$ at infinity. (*b*) At what distance from the sphere's surface has the electric potential decreased by 550 V?

28-10 The Electrostatic Accelerator

49. (*a*) How much charge is required to raise an isolated metallic sphere of 1.0-m radius to a potential of 1.0 MV? Assume $V = 0$ at infinity. Repeat for a sphere of 1.0-cm radius. (*b*) Why use a large sphere in an electrostatic accelerator when the same potential can be achieved using a smaller charge with a small sphere? (Hint: Calculate the charge densities.)

50. Let the potential difference between the high-potential inner shell of a Van de Graaff accelerator and the point at which charges are sprayed onto the moving belt be 3.41 MV. If the belt transfers charge to the shell at the rate of 2.83 mC/s, what minimum power must be provided to drive the belt?

PROBLEMS

1. (*a*) Through what potential difference must an electron fall, according to Newtonian mechanics, to acquire a speed v equal to the speed c of light? (*b*) Newtonian mechanics fails as $v \rightarrow c$. Therefore, using the correct relativistic expression for the kinetic energy (see Eq. 20-27)

$$K = mc^2\left[\frac{1}{\sqrt{1-(v/c)^2}} - 1\right]$$

in place of the Newtonian expression $K = \frac{1}{2}mv^2$, determine the actual electron speed acquired in falling through the potential difference computed in (*a*). Express this speed as an appropriate fraction of the speed of light.

2. Repeat Exercise 3, assuming that the electron is a hollow shell of radius 2.82×10^{-15} m with charge e evenly distributed over the surface.

3. A particle of (positive) charge Q is assumed to have a fixed position at P. A second particle of mass m and (negative) charge $-q$ moves at constant speed in a circle of radius r_1 centered at P. Derive an expression for the work W that must be done by an external agent on the second particle in order to increase the radius of the circle of motion, centered at P, to r_2.

4. The electric field inside a nonconducting sphere of radius R, containing uniform charge density, is radially directed and has magnitude

$$E = \frac{qr}{4\pi\epsilon_0 R^3},$$

where q is the total charge in the sphere and r is the distance from the center of the sphere. (*a*) Find the potential V inside the sphere, taking $V = 0$ at $r = 0$. (*b*) What is the difference in electric potential between a point on the surface and the center of the sphere? If q is positive, which point is at the higher potential? (*c*) Show that the potential at a distance r from the center, where $r < R$, is given by

$$V = \frac{q(3R^2 - r^2)}{8\pi\epsilon_0 R^3},$$

where the zero of potential is taken at $r = \infty$. Why does this result differ from that of part (*a*)?

5. Three charges of + 122 mC each are placed on the corners of an equilateral triangle, 1.72 m on a side. If energy is supplied at the rate of 831 W, how many days would be required to move one of the charges onto the midpoint of the line joining the other two?

6. A particle of mass m, charge $q > 0$, and initial kinetic energy K is projected (from an infinite separation) toward a heavy nucleus of charge Q, assumed to have a fixed position in our reference frame. (*a*) If the aim is "perfect," how close to the center of the nucleus is the particle when it comes instantaneously to rest? (*b*) With a particular imperfect aim, the particle's closest approach to the nucleus is twice the distance determined in part (*a*). Determine the speed of the particle at this closest distance of approach. Assume that the particle does not reach the surface of the nucleus.

7. A spherical drop of water carrying a charge of 32.0 pC has a potential of 512 V at its surface. (*a*) What is the radius of the drop? (*b*) If two such drops of the same charge and radius combine to form a single spherical drop, what is the potential at the surface of the new drop? Set $V = 0$ at infinity.

8. Figure 28-42 shows, edge-on, an "infinite" sheet of positive charge density σ. (*a*) How much work is done by the electric field of the sheet as a small positive test charge q_0 is moved from an initial position on the sheet to a final position located a perpendicular distance z from the sheet? (*b*) Use the result from (*a*) to show that the electric potential of an infinite sheet of charge can be written

$$V = V_0 - (\sigma/2\epsilon_0)z,$$

where V_0 is the potential at the surface of the sheet.

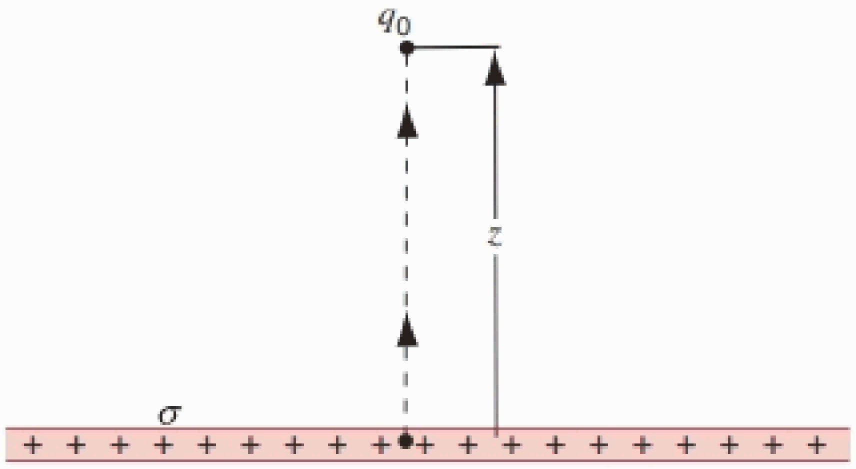


FIGURE 28-42. Problem 8.

9. A point charge $q_1 = +6e$ is fixed at the origin of a rectangular coordinate system, and a second point charge $q_2 = -10e$ is fixed at $x = 9.60$ nm, $y = 0$. With $V = 0$ at infinity, the locus of all points in the xy plane with $V = 0$ is a circle centered on the x axis, as shown in Fig. 28-43. Find (*a*) the location x_c of the center of the circle and (*b*) the radius R of the circle. (*c*) Is the $V = 5$ V equipotential also a circle?

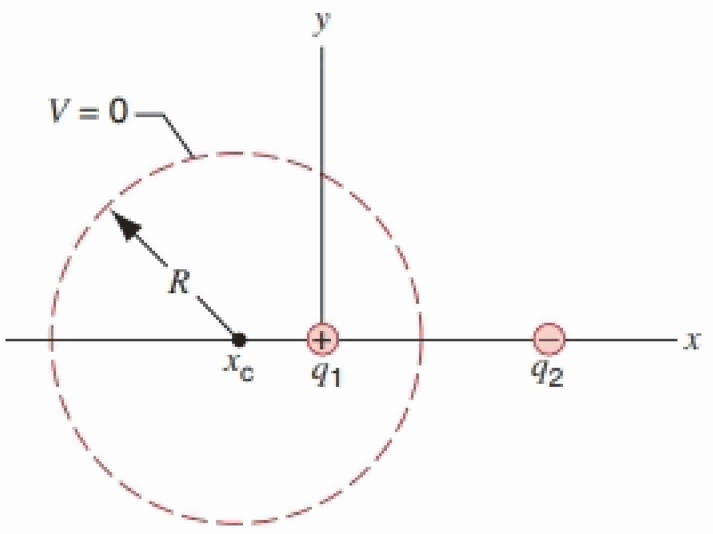


FIGURE 28-43. Problem 9.

10. A total amount of positive charge Q is spread onto a nonconducting, flat, circular annulus of inner radius a and outer radius b. The charge is distributed so that the charge density (charge per unit area) is given by $\sigma = k/r^3$, where r is the distance from the center of the annulus to any point on it. Show that (with $V = 0$ at infinity) the potential at the center of the annulus is given by

$$V = \frac{Q}{8\pi\epsilon_0}\left(\frac{a+b}{ab}\right).$$

11. For the charge configuration of Fig. 28-44, show that $V(r)$ for points on the vertical axis, assuming $r \gg d$, is given by

$$V = \frac{1}{4\pi\epsilon_0}\frac{q}{r}\left(1 + \frac{2d}{r}\right).$$

(Hint: The charge configuration can be viewed as the sum of an isolated charge and a dipole.) Set $V = 0$ at infinity.

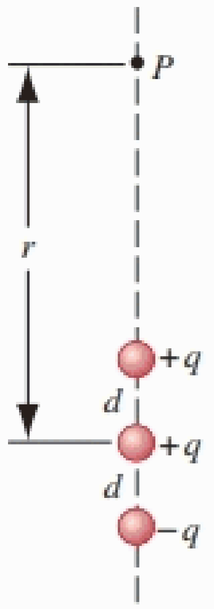


FIGURE 28-44. Problem 11.

12. A charge per unit length λ is distributed uniformly along a thin rod of length L. (*a*) Determine the potential (chosen to be zero at infinity) at a point P a distance y from one end of the rod and in line with it (see Fig. 28-45). (*b*) Use the result of (*a*) to compute the component of the electric field at P in the y direction (along the rod). (*c*) Determine the component of the electric field at P in a direction perpendicular to the rod.

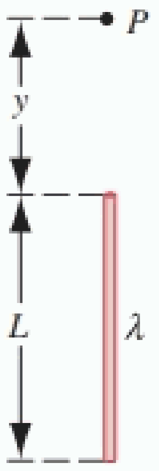


FIGURE 28-45. Problem 12.

13. On a thin rod of length L lying along the x axis with one end at the origin ($x = 0$), as in Fig. 28-46, there is distributed a charge per unit length given by $\lambda = kr$, where k is a constant and r is the distance from the origin. (*a*) Taking the electrostatic potential at infinity to be zero, find V at the point P on the y axis. (*b*) Determine the vertical component, E_y, of the electric field at P from the result of part (*a*) and also by direct calculation. (*c*) Why cannot E_x, the horizontal component of the electric field at P, be found using the result of part (*a*)? (*d*) At what distance from the rod along the y axis is the potential equal to one-half the value at the left end of the rod?

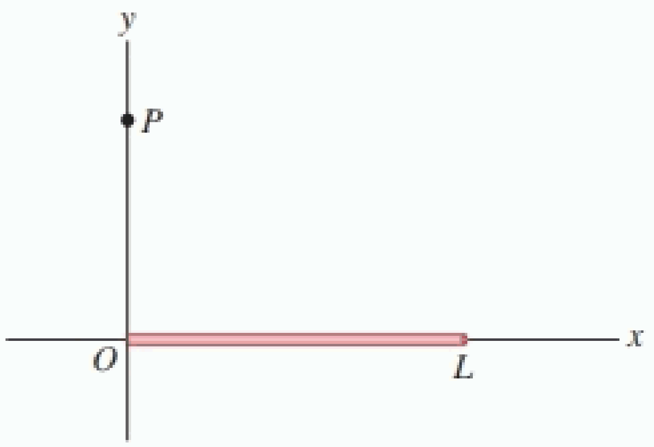


FIGURE 28-46. Problem 13.

14. Two identical conducting spheres of radius 15.0 cm are separated by a distance of 10.0 m. What is the charge on each sphere if the potential of one is $+1500$ V and the other is -1500 V? What assumptions have you made? Take $V = 0$ at infinity.

15. Consider the Earth to be a spherical conductor of radius 6370 km and to be initially uncharged. A metal sphere, having a radius of 13 cm and carrying a charge of -6.2 nC is earthed—that is, put into electrical contact with the Earth. Show that this process effectively discharges the sphere, by calculating the fraction of the excess electrons originally present on the sphere that remain after the sphere is earthed.

16. A copper sphere whose radius is 1.08 cm has a very thin surface coating of nickel. Some of the nickel atoms are radioactive, each atom emitting an electron as it decays. Half of these electrons enter the copper sphere, each depositing 100 keV of energy there. The other half of the electrons escape, each carrying away a charge of $-e$. The nickel coating has an activity of 10.0 mCi (= 10.0 millicuries = 3.70×10^8 radioactive decays per second). The sphere is hung from a long, nonconducting string and insulated from its surroundings. How long will it take for the potential of the sphere to increase by 1000 V?

17. Consider a thin, isolated, conducting, spherical shell that is uniformly charged to a constant charge density σ. How much work does it take to move a small positive test charge q_0 (*a*) from the surface of the shell to the interior, through a small hole; (*b*) from one point on the surface to another, regardless of path; (*c*) from point to point inside the shell; and (*d*) from any point P outside the shell over any path, whether or not it pierces the shell, back to P? (*e*) For the conditions given, does it matter whether or not the shell is conducting?

18. The high-voltage electrode of an electrostatic accelerator is a charged spherical metal shell having a potential $V = +9.15$ MV (relative to $V = 0$ at infinity). (*a*) Electrical breakdown occurs in the gas in this machine at a field $E = 100$ MV/m. To prevent such breakdown, what restriction must be made on the radius r of the shell? (*b*) A long, moving, rubber belt transfers charge to the shell at 320 μC/S, the potential of the shell remaining constant because of leakage. What minimum power is required to transfer the charge? (*c*) The belt is of width $w = 48.5$ cm and travels at speed $v = 33.0$ m/s. What is the surface charge density on the belt?

COMPUTER PROBLEMS

1. The charge density on a rod of length L centered on the x axis is given by $\lambda = (1.0\ \mu\text{C/m}) \sin^2(\pi x/L)$. (*a*) Numerically generate a plot of the potential in the xy plane, and then use your plot to generate equipotential lines. (*b*) From the plot, generate electric field lines and compare your result to Computer Problem 2 in Chapter 26.

2. Numerically verify that in two dimensions the equipotential lines around two equal but opposite charges are circles. Are the circles concentric?

CHAPTER 29

THE ELECTRICAL PROPERTIES OF MATERIALS

Although ordinary matter is electrically neutral, containing equal numbers of positive and negative charges, materials can reveal a great range of different behaviors when they are placed in electric fields. Some materials can conduct electricity even in very small fields, whereas others remain nonconducting in enormously large fields. In some materials that do not permit the movement of charge, the electrical properties are determined by the rotation of dipoles in an applied field, yet in others the applied field can create dipoles where they did not exist before.

In this chapter we shall consider the basic behavior of two kinds of materials: conductors and insulators. We will show how we can understand their behavior in applied fields based on simple models of forces and movement of charges. Even though a detailed understanding of the electrical properties of materials requires the methods of quantum mechanics, we can learn a great deal about materials from classical models that ignore the quantum behavior.

29-1 TYPES OF MATERIALS

Natural and artificially made materials show a wide range of electrical properties. These properties are determined partly by the behavior of individual atoms or molecules and partly by the interactions of atoms or molecules in the bulk material. The ability of a material to conduct electricity may also depend on the conditions of the material, such as its temperature and pressure.

Conductors (for example, most metals) are materials in which electric charges readily flow. In many metals, each atom gives up one or more of its outer or valence electrons to the entire material, and we often regard the electrons as forming a "gas" within the material rather than belonging to any particular atom. These electrons are free to move when an electric field is applied to the material. Under static conditions the electric field in the interior of a conductor is zero, even if the conductor carries a net charge. (If this were not so, the free electrons would be accelerated, which would violate the assumption of a static charge distribution.) In Section 29-2, we will discuss the effect of an external electric field applied to a conductor under static conditions.

In an *insulator,* on the other hand, the electrons are bound rather tightly to the atoms and are not free to move under the electric fields that might be applied under ordinary circumstances. An insulator can carry any distribution of electric charges on its surface or in its interior, and (in contrast to a conductor) the electric field in the interior of an insulator can have nonzero values.

An insulating material can often be regarded as a collection of molecules that are not easily ionized. In this case the electrical properties may depend on the electric dipole moment of the molecules. Materials in which the molecules have permanent dipole moments are called *polar,* and electric fields can align the dipole moments of molecules, as we discussed in Section 26-7. In some materials, the alignment of the dipoles remains even when we remove the applied field; these materials are called *ferroelectric* (in analogy with ferromagnetic materials, in which *magnetic* dipole moments remain aligned even when an external *magnetic* field is removed). Even nonpolar materials can show these effects, because the applied electric field can induce a dipole moment in the molecules. These effects are discussed in Section 29-5.

Ordinary matter is usually electrically neutral. In the absence of an external electric field, this neutrality applies to individual atoms as well as to the entire material. The application of an electric field can remove one or more electrons from atoms of the material. This process is called *ionization,* and the resulting positively charged atoms with a deficiency of electrons are called *ions.* In an insulator, a sufficiently large electric field can ionize the atoms, and as a result there are electrons available to move through the material. Under these circumstances an insulator can behave more like a conductor. This situation is called *breakdown* and requires fields typically in the range of 10^6 V/m in air to 10^7 V/m in plastics and ceramics.

Intermediate between insulators and conductors are *semiconductors.* In a semiconductor, perhaps one atom in 10^{10} to 10^{12} might contribute an electron to the flow of electricity in the material (in contrast to a conductor, in which *every* atom typically contributes an electron to the flow of electricity). Commonly used semiconductors include silicon and germanium, as well as many compounds.

Even the best conductors (copper, silver, and gold) show a small but definitely nonzero resistance to the flow of electricity. Under certain conditions, often involving cooling to very low temperatures, electric charge can flow through some materials with no resistance at all. This property of materials is called *superconductivity,* and the materials under these conditions are called *superconductors.* Some materials can be relatively poor conductors at room temperature but can be superconductors at low temperatures.

In this chapter we study ways in which conductors and insulators respond to applied electric fields. Understanding the behavior of semiconductors and superconductors requires the methods of quantum mechanics, which are discussed in Chapter 49.

29-2 A CONDUCTOR IN AN ELECTRIC FIELD: STATIC CONDITIONS

Suppose we place a large rectangular slab of a conductor such as copper in a uniform electric field, as shown in Fig. 29-1*a*. We can regard the copper as a "gas" of electrons that are free to move in a lattice of copper ions in fixed locations. The electric field $\vec{\mathbf{E}}_0$ exerts a force $\vec{\mathbf{F}} = -e\vec{\mathbf{E}}_0$ on the electrons, which causes the electrons to move in a direction opposite to the field. The electrons quickly move to the top surface of the copper, leaving a deficiency of electrons (a positive charge) on the bottom surface. When we place a conductor in an external field, the charges redistribute themselves almost instantaneously, after which electrostatic conditions apply.

The two surfaces of the conductor can be considered as sheets of charge, which set up an electric field $\vec{\mathbf{E}}'$ as shown in Fig. 29-1*b*. Inside the copper, the net electric field $\vec{\mathbf{E}}$ is the vector sum of the two fields: $\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 + \vec{\mathbf{E}}'$. In terms of magnitudes, the sum becomes a difference, because the two fields are in opposite directions: $E = E_0 - E'$. In the interior of the copper under static conditions, the net electric field E must be zero, as we discussed in Section 27-6. (In Section 27-6, we did not consider the presence of an externally applied electric field; however, the conclusion remains the same—the electric field inside the conductor must be zero, because otherwise the free electrons in the conductor would be accelerated, thus violating our assumption of a static situation.) The applied electric field E_0 must move just enough electrons to the surface to set up an electric field E' that has the same magnitude as E_0, giving a net field of zero inside the copper (Fig. 29-1*c*). Outside the slab, the sheets of charge on the two surfaces give electric fields that cancel, leaving the net field unchanged in those regions.

Figure 29-2 shows an uncharged conductor of irregular shape in an originally uniform electric field. Once again, free electrons in the conductor quickly move to the surface,

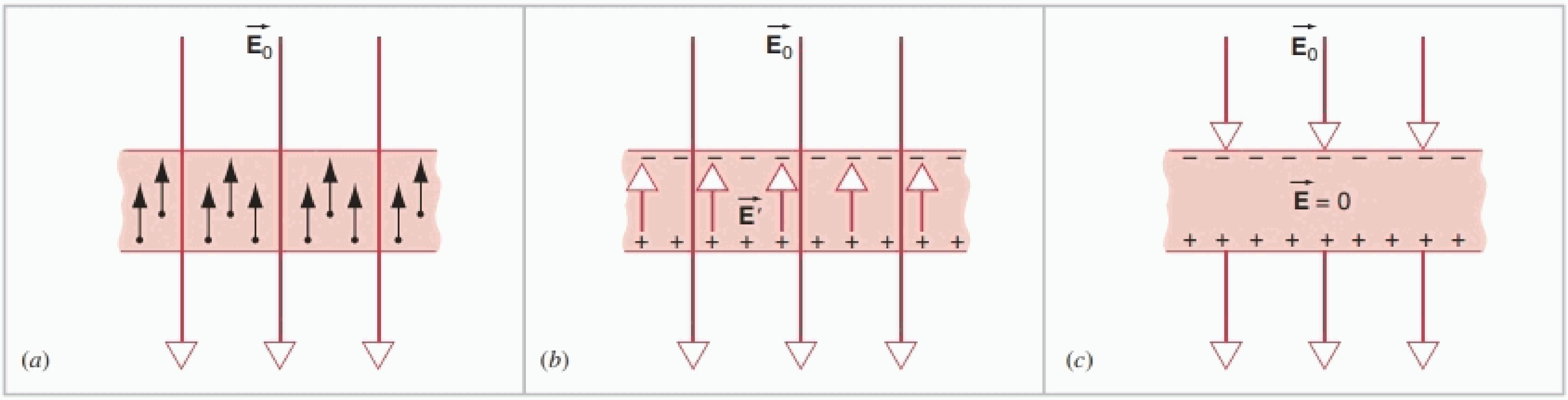


FIGURE 29-1. (*a*) A large slab of conductor is placed in a uniform electric field. Electrons in the material move upward in response to the field. (*b*) Electrons accumulate on the top surface, leaving positive ions on the bottom. These charges set up a field $\vec{\mathbf{E}}'$. (*c*) Inside the slab, the net field is zero.

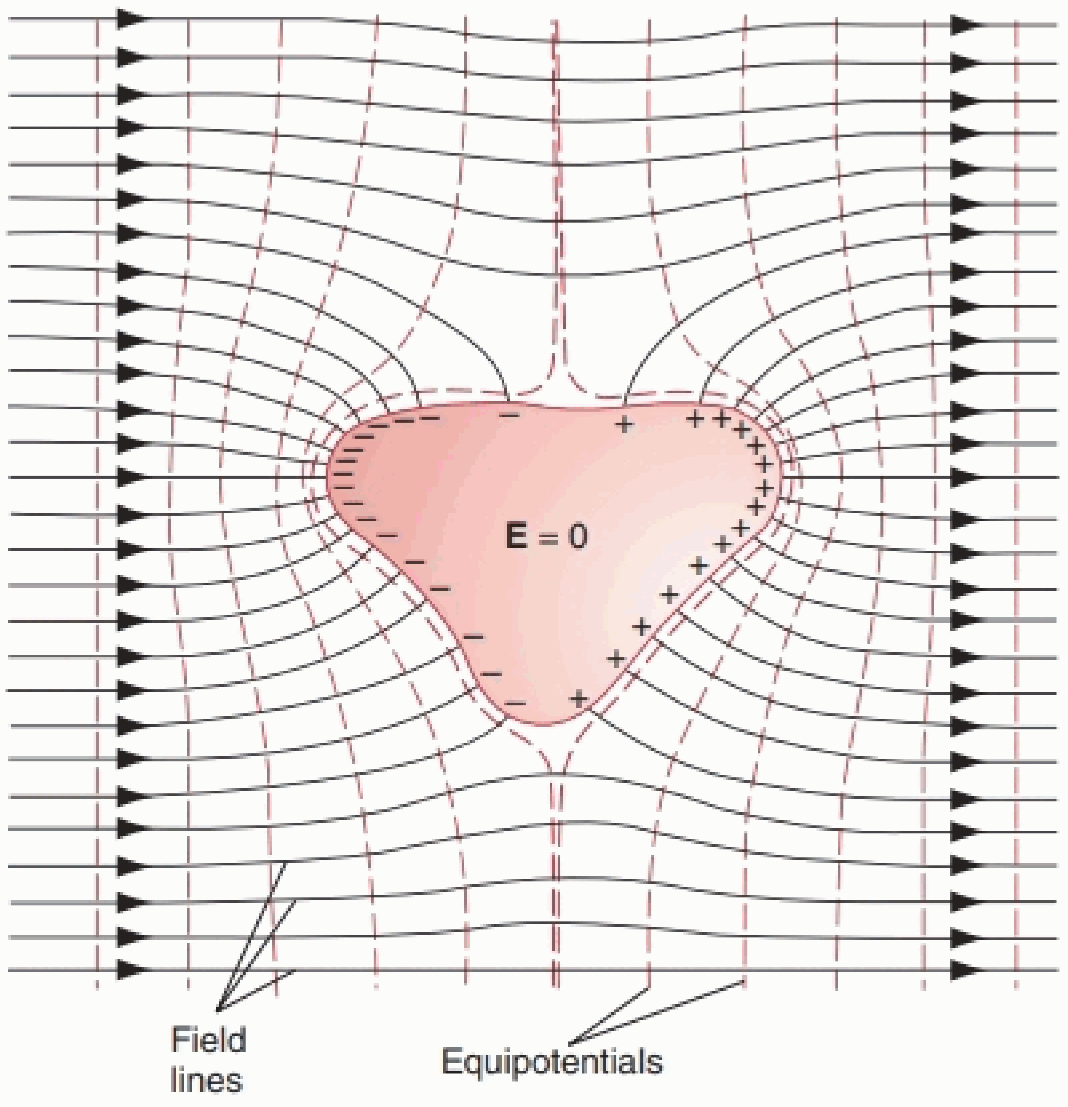


FIGURE 29-2. An uncharged conductor is placed in an external electric field. The conduction electrons distribute themselves on the surface to produce a charge distribution as shown, reducing the field inside the conductor to zero. Note the distortion of the lines of force (solid lines) and the equipotentials (dashed lines) when the conductor is placed in the previously uniform field. Note also how the electric field lines originate on positive charges and terminate on negative charges.

establishing a distribution of positive and negative charges that gives an electric field exactly canceling the applied field in the interior of the conductor. Outside the conductor, the field is the (vector) sum of the original uniform field and the field due to the charges on the surface of the conductor. Note that the field lines originate on positive charges and terminate on negative charges. Note also that the charge density is large on parts of the surface where the radius of curvature is small, as we discussed in Section 28-9, and that the field is large (the field lines are close together) where the charge density is large.

At the surface of the conductor in Fig. 29-2, the electric field lines are perpendicular to the surface. If this were not true, then there would be a component of the electric field parallel to the surface, which would cause charges to move. Since this would violate our assumption of a static situation, this component of the electric field cannot exist and the field must be perpendicular to the surface.

The figure also shows the equipotentials for this situation. Far from the conductor, where the field is uniform, the equipotentials are flat planes. As we move close to the conductor, the equipotentials are distorted, until at the surface the equipotential follows the surface exactly; as we discussed in Section 28-9, the surface of a conductor is an equipotential.

SAMPLE PROBLEM 29-1. A large, thin plate of copper is placed in a uniform electric field of magnitude $E_0 = 450$ N/C that is perpendicular to the plate (as in Fig. 29-1). Find the resulting surface charge density on the copper.

Solution The electric field causes a positive charge density σ on the lower surface of the plate and a negative charge density of equal magnitude on the upper surface. The field in the interior of the plate must be zero, which means that the two charge distributions must combine to give an electric field inside the plate of magnitude E_0 and direction opposite to the applied field. If we regard the plate as of very large dimensions, the field due to the positive charge distribution is, according to Eq. 26-20, $E_+ = \sigma/2\epsilon_0$, and the magnitude of the field due to the negative charge is $E_- = \sigma/2\epsilon_0$. These two fields are in the same direction and must add to give a total field of E_0:

$$E_0 = \sigma/2\epsilon_0 + \sigma/2\epsilon_0 = \sigma/\epsilon_0$$

and the charge density on each surface is

$$\begin{aligned}\sigma = \epsilon_0 E_0 &= (8.85 \times 10^{-12}\ \mathrm{C^2/N\cdot m^2})(450\ \mathrm{N/C})\\ &= 3.98 \times 10^{-9}\ \mathrm{C/m^2}.\end{aligned}$$

Note that outside the copper plate, the fields due to the two sheets of charge cancel one another, so that the resulting field remains E_0. This is true only for the flat geometry of this problem and is not true in general; see for example Fig. 29-2.

29-3 A CONDUCTOR IN AN ELECTRIC FIELD: DYNAMIC CONDITIONS

In Fig. 29-1*a*, electrons move from the bottom of the slab of copper to the top under the action of the applied electric field, until the concentration of electrons at the top of the slab (and of positive ions at the bottom) creates a field that cancels the applied field in the interior of the copper and prevents the flow of additional electrons. Suppose there were a mechanism to remove electrons from the top of the slab, carry them around an external path, and re-inject them at the bottom of the slab (shown schematically in Fig. 29-3). In this case, there would be no build-up of charge on the top and bottom of the slab, and the electrostatic conditions of the previous section cannot be applied to the copper. In particular, the conclusion drawn in the last section is no longer valid—the electric field inside the copper will in general be nonzero when charges are flowing.

The continuous loop of flowing electrons is a simple representation of an electric circuit, and the flow of electrons (or other charged particles) is called an *electric current.*

Let us examine the flow of electric charge past a particular point in the interior of the material (Fig. 29-4). A quantity of charge dq will pass through a small surface of area A in a time dt. For example, the area A might be the cross-sectional area of a wire through which the charge

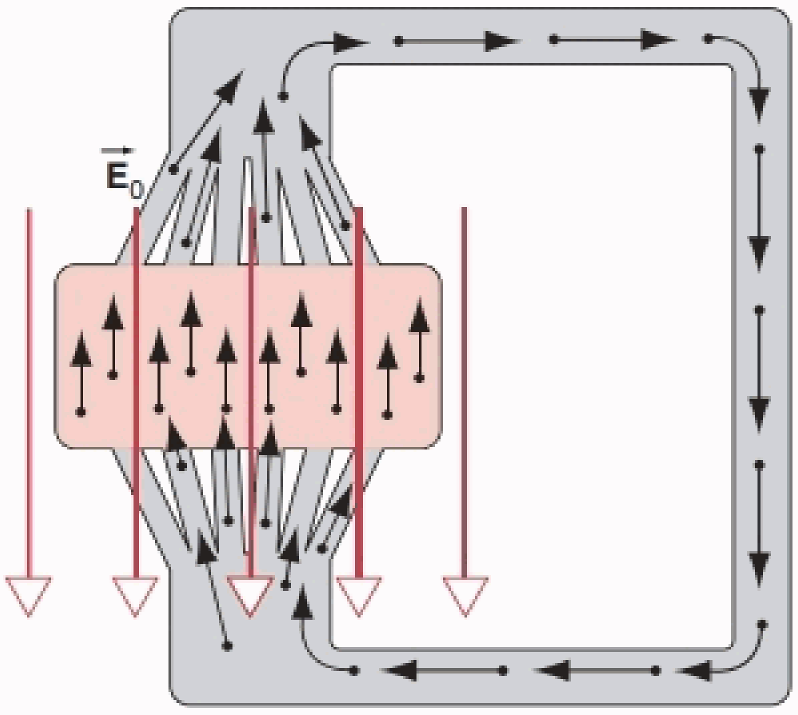


FIGURE 29-3. The electric field $\vec{\mathbf{E}}_0$ moves electrons through the slab of copper. The electrons can be collected at the top of the slab and transported through an external path to the bottom of the copper slab.

is flowing. The *electric current i* is defined as the net charge that flows through the surface per unit time interval:

$$i = dq/dt. \tag{29-1}$$

For electric current to exist, there must be a net flow of charge across the surface. If neutral atoms travel across the surface, no current is flowing even though charges travel across the surface, because equal numbers of positive and negative charges cross the surface. If electrons are traveling randomly through the material, with equal numbers crossing the surface in either direction, no current flows because the net charge crossing the surface is zero.

The electric current has a direction, which is defined to be the direction of the flow of positive charge. Even though the current has a direction, current is a scalar and not a vector, because currents do not satisfy the laws for vector addition.

The SI unit of current is the *ampere* (A), defined as

$$1 \text{ ampere} = 1 \text{ coulomb/second}.$$

If the current is constant, then Eq. 29-1 becomes

$$i = q/t. \tag{29-2}$$

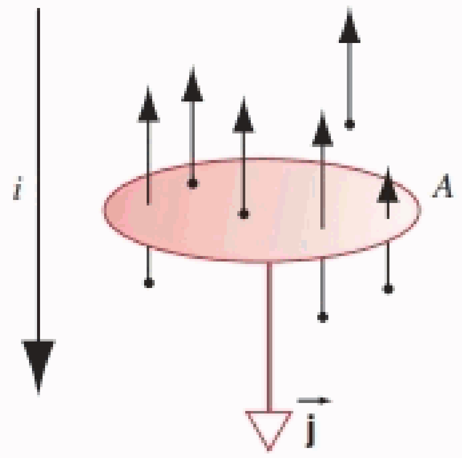


FIGURE 29-4. Electrons pass through an area A. The directions of the current i and of the vector current density $\vec{\mathbf{j}}$ are opposite to the motion of the electrons.

The net charge passing through any surface is determined by integrating the current:

$$q = \int i\, dt. \tag{29-3}$$

A related vector quantity is the *current density* $\vec{\mathbf{j}}$ or current per unit area, whose magnitude is defined as

$$j = i/A. \tag{29-4}$$

The direction of $\vec{\mathbf{j}}$ is defined to be the direction of the flow of positive charge. Since electrons are moving upward in Fig. 29-4, the direction of $\vec{\mathbf{j}}$ is downward. That is, electrons move in the direction of $-\vec{\mathbf{j}}$.

The current passing through any surface can be determined by integrating the current density over the surface:

$$i = \int \vec{\mathbf{j}} \cdot d\vec{\mathbf{A}}, \tag{29-5}$$

where $d\vec{\mathbf{A}}$ is an element of surface area and the integral is done over the entire surface through which we want to find the current. The vector $d\vec{\mathbf{A}}$ is taken to be perpendicular to the surface element such that $\vec{\mathbf{j}} \cdot d\vec{\mathbf{A}}$ is positive, corresponding to a positive current i.

Current Density and Drift Speed

As electrons make their way through the copper, they are accelerated by an electric field, which exerts a force $-e\vec{\mathbf{E}}$ on the electrons. In Section 29-2 we considered static conditions, in which the electric field is always zero inside a conductor. Here we consider charges in motion, so that static conditions do not apply and $\vec{\mathbf{E}}$ can be nonzero inside a conductor.

The electrons collide with the ions of the lattice and transfer energy to them. The motion of individual electrons is therefore very irregular, consisting of a short interval of acceleration in a direction opposite to the electric field, followed by a collision with an ion that might send the electron into motion in any direction, followed by another acceleration, and so on. The net effect is a drift of electrons in a direction opposite to the field. There is no net acceleration of electrons, because they continually lose energy in collisions with the lattice of copper ions. In effect, energy is transferred from the applied field to the lattice (in the form of internal energy of the conductor, often observed as a temperature increase). On the average, electrons can be described as moving with a constant *drift velocity* $\vec{\mathbf{v}}_\mathrm{d}$ in a direction opposite to the field, as indicated in Fig. 29-5.

Consider the motion of electrons in a portion of the conductor of length L. The electrons are moving with drift speed v_d, so they travel the length L in a time $t = L/v_\mathrm{d}$. The conductor has a cross-sectional area A, so in the time t all of the electrons in the volume AL will travel through a surface at the right end of the conductor. If the density of electrons (number per unit volume) is n, then the magnitude of

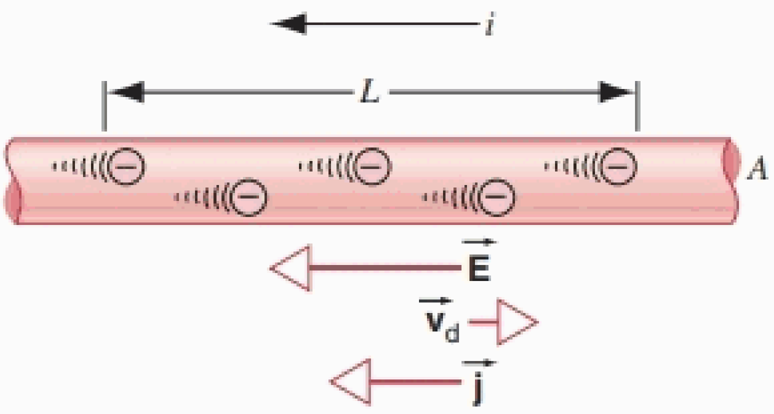


FIGURE 29-5. The electric field causes electrons to drift to the right. The conventional current (the hypothetical direction of flow of positive charge) is to the left. The current density $\vec{j}$ is likewise drawn as if the charge carriers were positive, so that $\vec{j}$ and $\vec{E}$ are in the same direction.

the net charge passing through the surface is $q = enAL$, and the current density is

$$j = \frac{q}{At} = \frac{enAL}{AL/v_d} = env_d. \tag{29-6}$$

In vector notation, this is

$$\vec{j} = -en\vec{v}_d. \tag{29-7}$$

The negative sign again reminds us that the direction of the current density is opposite to the motion of the electrons.

As the following sample problems illustrate, the drift speeds of electrons in typical materials are very small compared with the speed of the random thermal motions of electrons (typically 10^6 m/s).

SAMPLE PROBLEM 29-2. One end of an aluminum wire whose diameter is 2.5 mm is welded to one end of a copper wire whose diameter is 1.8 mm. The composite wire carries a steady current i of 1.3 A. What is the current density in each wire?

Solution We may take the current density as (a different) constant within each wire except for points near the junction. The current density is given by Eq. 29-4, $j = i/A$. The cross-sectional area A of the aluminum wire is

$$A_{\mathrm{Al}} = \tfrac{1}{4}\pi d^2 = (\pi/4)(2.5 \times 10^{-3}\ \mathrm{m})^2 = 4.91 \times 10^{-6}\ \mathrm{m}^2,$$

so that

$$j_{\mathrm{Al}} = \frac{1.3\ \mathrm{A}}{4.91 \times 10^{-6}\ \mathrm{m}^2} = 2.6 \times 10^5\ \mathrm{A/m}^2 = 26\ \mathrm{A/cm}^2.$$

As you can verify, the cross-sectional area of the copper wire is $2.54 \times 10^{-6}\ \mathrm{m}^2$, so that

$$j_{\mathrm{Cu}} = \frac{1.3\ \mathrm{A}}{2.54 \times j10^{-6}\ \mathrm{m}^2} = 5.1 \times 10^5\ \mathrm{A/m}^2 = 51\ \mathrm{A/cm}^2.$$

The fact that the wires are of different materials does not enter here.

SAMPLE PROBLEM 29-3. What is the drift speed of the conduction electrons in the copper wire of Sample Problem 29-2?

Solution In copper, there is very nearly one conduction electron per atom on the average. The number n of electrons per unit volume is therefore the same as the number of atoms per unit volume and is determined from

$$\frac{n}{N_A} = \frac{\rho_m}{M} \quad \text{or} \quad \frac{\mathrm{atoms/m^3}}{\mathrm{atoms/mol}} = \frac{\mathrm{mass/m^3}}{\mathrm{mass/mol}}.$$

Here ρ_m is the (mass) density of copper, N_A is the Avogadro constant, and M is the molar mass of copper.* Thus

$$n = \frac{N_A\rho_m}{M} = \frac{(6.02 \times 10^{23}\ \mathrm{electrons/mol})(8.96 \times 10^3\ \mathrm{kg/m^3})}{63.5 \times 10^{-3}\ \mathrm{kg/mol}}$$

$$= 8.49 \times 10^{28}\ \mathrm{electrons/m^3}.$$

We then have, using Eq. 29-6 ($v_d = j/ne$),

$$v_d = \frac{5.1 \times 10^5\ \mathrm{A/m^2}}{(8.49 \times 10^{28}\ \mathrm{electrons/m^3})(1.60 \times 10^{-19}\ \mathrm{C/electron})}$$

$$= 3.8 \times 10^{-5}\ \mathrm{m/s} = 14\ \mathrm{cm/h}.$$

You should be able to show that for the aluminum wire, $v_d = 2.7 \times 10^{-5}$ m/s = 9.7 cm/h. Can you explain, in physical terms, why in this example the drift speed is smaller in aluminum than in copper, even though the two wires carry the same current?

If the electrons drift at such a low speed, why do electrical effects seem to occur immediately after a switch is thrown, such as when you turn on the room lights? Confusion on this point results from not distinguishing between the drift speed of the electrons and the speed at which *changes* in the electric field configuration travel along wires. This latter speed approaches that of light. Similarly, when you turn the valve on your garden hose, with the hose full of water, a pressure wave travels along the hose at the speed of sound in water. The speed at which the water moves through the hose—measured perhaps with a dye marker—is much lower.

SAMPLE PROBLEM 29-4. A strip of silicon, of cross-sectional width $w = 3.2$ mm and thickness $d = 250\ \mu$m, carries a current i of 190 mA. The silicon is an *n-type semiconductor,* having been "doped" with a controlled amount of phosphorus impurity. The doping has the effect of greatly increasing n, the number of charge carriers (electrons, in this case) per unit volume, as compared with the value for pure silicon. In this case, $n = 8.0 \times 10^{21}\ \mathrm{m}^{-3}$. (*a*) What is the current density in the strip? (*b*) What is the drift speed?

Solution (*a*) From Eq. 29-4,

$$j = \frac{i}{wd} = \frac{190 \times 10^{-3}\ \mathrm{A}}{(3.2 \times 10^{-3}\ \mathrm{m})(250 \times 10^{-6}\ \mathrm{m})}$$

$$= 2.4 \times 10^5\ \mathrm{A/m^2}.$$

(*b*) From Eq. 29-6,

$$v_d = \frac{j}{ne} = \frac{2.4 \times 10^5\ \mathrm{A/m^2}}{(8.0 \times 10^{21}\ \mathrm{m^{-3}})(1.60 \times 10^{-19}\ \mathrm{C})} = 190\ \mathrm{m/s}.$$

The drift speed (190 m/s) of the electrons in this silicon semiconductor is much greater than the drift speed (3.8×10^{-5} m/s) of the conduction electrons in the metallic copper conductor of Sample Problem 29-3, even though the current densities are similar. The number of charge carriers in this semiconductor (8.0 ×

---

*We use the subscript m to make it clear that the density referred to here is a mass density (kg/m³), not a charge density (C/m³).

10^{21} m^{-3}) is much smaller than the number of charge carriers in the copper conductor (8.49×10^{28} m^{-3}). The smaller number of charge carriers must drift faster in the semiconductor if they are to establish the same current density that the greater number of charge carriers establish in copper.

29-4 OHMIC MATERIALS

Between collisions with the lattice ions, the electrons in a conducting material are accelerated by the electric field $\vec{\mathbf{E}}$, and so their drift velocity is proportional to $\vec{\mathbf{E}}$. The current density $\vec{\mathbf{j}}$ is also proportional to $\vec{\mathbf{v}}_d$, so it is reasonable that $\vec{\mathbf{j}}$ should be proportional to $\vec{\mathbf{E}}$. In fact, we observe this type of behavior for a wide class of materials. The proportionality constant between the current density and electric field is the *electrical conductivity* σ of the material:

$$\vec{\mathbf{j}} = \sigma\vec{\mathbf{E}}. \tag{29-8}$$

A large value of σ indicates that the material is a good conductor of electric current. The conductivity is a property of the material, not of any particular sample of the material. The SI unit for conductivity is the *siemens per meter* (S/m), where the siemens is defined as

$$1 \text{ siemens} = 1 \text{ ampere/volt}.$$

It is more common to find materials characterized by their *resistivity,* which is the inverse of the conductivity:

$$\rho = 1/\sigma, \tag{29-9}$$

in which case Eq. 29-8 becomes

$$\vec{\mathbf{E}} = \rho\vec{\mathbf{j}}. \tag{29-10}$$

Units of resistivity are *ohm · meter,* where the ohm (symbol Ω) is defined as

$$1 \text{ ohm} = 1 \text{ volt/ampere}.$$

Note that 1 ohm = (1 siemens)$^{-1}$.

Equations 29-8 and 29-10 are valid only for isotropic materials, whose electrical properties are the same in all directions. In these materials, $\vec{\mathbf{j}}$ will always be in the same direction as $\vec{\mathbf{E}}$.

Table 29-1 gives some values of the resistivity for various materials. A perfect insulator would have $\rho = \infty$ (or $\sigma = 0$). Note that even good insulators are weakly conducting.

We can use Eq. 29-10 to determine the resistivity of any material by applying an electric field and measuring the resulting current density. For some materials, we find that the resistivity is not a constant but depends on the strength of the electric field. That is, if we double the electric field the current density does not double. For other materials, we find that the resistivity does not depend on the strength of the applied field for a wide range of applied fields. For such materials, a plot of E against j gives a straight line, whose slope is the resistivity ρ. These materials are known as *ohmic* materials. Equivalently, such materials are said to satisfy *Ohm's law:*

TABLE 29-1 **Resistivity of Some Materials at Room Temperature (20°C)**

Material	*Resistivity,* ρ ($\Omega \cdot$ m)	*Temperature Coefficient of Resistivity* α_{av} (per C°)
Typical Metals		
Silver	1.62×10^{-8}	4.1×10^{-3}
Copper	1.69×10^{-8}	4.3×10^{-3}
Aluminum	2.75×10^{-8}	4.4×10^{-3}
Tungsten	5.25×10^{-8}	4.5×10^{-3}
Iron	9.68×10^{-8}	6.5×10^{-3}
Platinum	10.6×10^{-8}	3.9×10^{-3}
Manganin[a]	48.2×10^{-8}	0.002×10^{-3}
Typical Semiconductors		
Silicon pure	2.5×10^{3}	-70×10^{-3}
Silicon *n*-type[b]	8.7×10^{-4}	
Silicon *p*-type[c]	2.8×10^{-3}	
Typical Insulators		
Pure water	2.5×10^{5}	
Glass	$10^{10} - 10^{14}$	
Polystyrene	$> 10^{14}$	
Fused quartz	$\approx 10^{16}$	

[a] An alloy specifically designed to have a small value of α.

[b] Pure silicon "doped" with phosphorus impurities to a charge carrier density of 10^{23} m^{-3}.

[c] Pure silicon "doped" with aluminum impurities to a charge carrier density of 10^{23} m^{-3}.

> *The resistivity (or conductivity) of a material is independent of the magnitude and direction of the applied electric field.*

Many homogeneous materials, including conducting metals such as copper, obey Ohm's law for a certain range of values of the applied electric field. If the field is sufficiently large, all materials will behave in violation of Ohm's law.

The resistivity values in Table 29-1 are properties of the materials listed. We might also want to know the *resistance* of a particular object, such as a block of copper of certain dimensions. Figure 29-6 illustrates the situation for a homogeneous, isotropic conductor of length L and uniform cross-sectional area A, to which we have applied a potential difference ΔV. Inside the object, there is a uniform electric

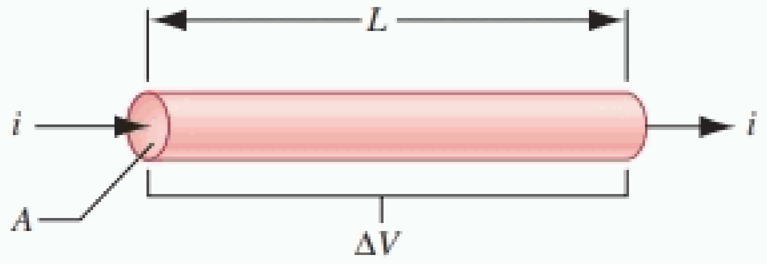


FIGURE 29-6. A potential difference ΔV is applied across a cylindrical conductor of length L and cross-sectional area A, establishing a current i.

field $E = \Delta V/L$. If the current density is also uniform over the area A, then $j = i/A$. The resistivity is then

$$\rho = \frac{E}{j} = \frac{\Delta V/L}{i/A}. \tag{29-11}$$

The quantity $\Delta V/i$ that appears in this equation is defined as the resistance R:

$$R = \frac{\Delta V}{i}. \tag{29-12}$$

Combining Eqs. 29-11 and 29-12, we obtain an expression for the resistance R:

$$R = \rho \frac{L}{A}. \tag{29-13}$$

The resistance R is characteristic of a particular object and depends on the material of which it is made as well as on its length and cross-sectional area; the resistivity ρ is characteristic of the material in general. The units of resistance are ohms (Ω).

Equation 29-12 gives us another basis for stating Ohm's law. For a particular object, we can measure the current i for various applied potential differences and plot i as a function of ΔV. If this plot gives a straight line, then the object is ohmic and obeys Ohm's law. An equivalent statement of Ohm's law is:

> *The resistance of an object is independent of the magnitude or sign of the applied potential difference.*

Ordinary resistors that are found in electric circuits are ohmic for the range of potential differences that are normally used in circuits. Semiconducting devices, such as diodes and transistors, usually are nonohmic. Figure 29-7 compares the current–voltage plots for ohmic and non-ohmic devices.

Keep in mind that the relationship $\Delta V = iR$ is *not* a statement of Ohm's law. This equation defines the resistance and is true for both ohmic and nonohmic objects. Even for nonohmic devices, we can find a value of the resistance R for a particular value of ΔV; for a different ΔV, a different value of R will be obtained. For ohmic devices, we get the same value of R for any value of ΔV.

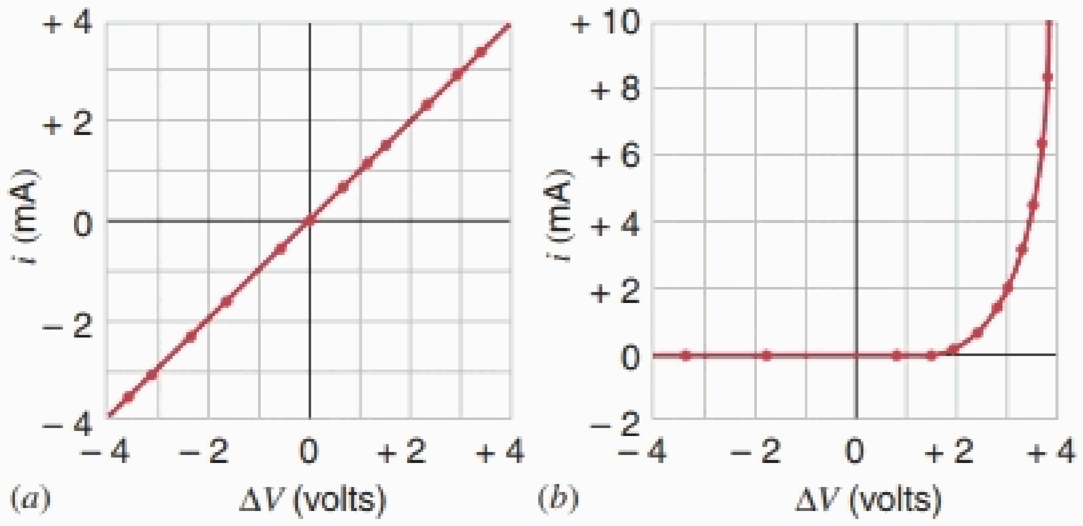


FIGURE 29-7. (*a*) A current–voltage plot for a material that obeys Ohm's law, in this case a 1000-Ω resistor. (*b*) A current–voltage plot for a material that does not obey Ohm's law, in this case a *pn* junction diode.

ΔV, i, and R are *macroscopic* quantities, applying to a particular body or extended region. The corresponding *microscopic* quantities are $\vec{\mathbf{E}}$, $\vec{\mathbf{j}}$, and ρ (or σ); they have values at every point in a body. The macroscopic quantities are related by Eq. 29-12 ($\Delta V = iR$) and the microscopic quantities by Eq. 29-10 ($\vec{\mathbf{E}} = \rho\vec{\mathbf{j}}$).

The macroscopic quantities ΔV, i, and R are of primary interest when we are making electrical measurements on real conducting objects. They are the quantities whose values are indicated on meters. The microscopic quantities $\vec{\mathbf{E}}$, $\vec{\mathbf{j}}$, and ρ are of primary importance when we are concerned with the fundamental behavior of matter (rather than of specimens of matter), as we usually are in the research area of *solid state* (or *condensed matter*) physics. Section 29-5 accordingly deals with an atomic view of the *resistivity* of a metal and not of the *resistance* of a metallic specimen.

SAMPLE PROBLEM 29-5. A rectangular block of iron has dimensions 1.2 cm × 1.2 cm × 15 cm. (*a*) What is the resistance of the block measured between the two square ends? (*b*) What is the resistance between two opposing rectangular faces? The resistivity of iron at room temperature is $9.68 \times 10^{-8}\ \Omega\cdot\text{m}$.

Solution (*a*) The area of a square end is $(1.2 \times 10^{-2}\ \text{m})^2$ or $1.44 \times 10^{-4}\ \text{m}^2$. From Eq. 29-13,

$$R = \frac{\rho L}{A} = \frac{(9.68 \times 10^{-8}\ \Omega\cdot\text{m})(0.15\ \text{m})}{1.44 \times 10^{-4}\ \text{m}^2}$$
$$= 1.0 \times 10^{-4}\ \Omega = 100\ \mu\Omega.$$

(*b*) The area of a rectangular face is $(1.2 \times 10^{-2}\ \text{m})(0.15\ \text{m})$ or $1.80 \times 10^{-3}\ \text{m}^2$. From Eq. 29-13,

$$R = \frac{\rho L}{A} = \frac{(9.68 \times 10^{-8}\ \Omega\cdot\text{m})(1.2 \times 10^{-2}\ \text{m})}{1.80 \times 10^{-3}\ \text{m}^2}$$
$$= 6.5 \times 10^{-7}\ \Omega = 0.65\ \mu\Omega.$$

We assume in each case that the potential difference is applied to the block in such a way that the surfaces between which the resistance is desired are equipotentials. Then the electric field is uniform between the surfaces, and as a result the current density is also uniform. Otherwise, Eq. 29-13 would not apply.

Analogy between Current and Heat Flow (Optional)

A close analogy exists between the flow of charge established by a potential difference and the flow of heat established by a temperature difference. Consider a thin electrically conducting slab of thickness Δx and area A. Let a potential difference ΔV be maintained between opposing faces. The current i is given by Eqs. 29-12 ($i = \Delta V/R$) and 29-13 ($R = \rho L/A$), or

$$i = \frac{\Delta V}{R} = \frac{\Delta V}{\rho \Delta x/A} = \sigma A \frac{\Delta V}{\Delta x}$$

using $\sigma = \rho^{-1}$. In the limiting case of a slab of thickness dx this becomes

$$\frac{dq}{dt} = -\sigma A \frac{dV}{dx}. \qquad (29\text{-}14)$$

The negative sign in Eq. 29-14 indicates that positive charge flows in the direction of decreasing V; that is, dq/dt is positive when dV/dx is negative.

The analogous heat flow equation (see Section 23-2) is

$$\frac{dQ}{dt} = -kA \frac{dT}{dx}, \qquad (29\text{-}15)$$

which shows that k, the thermal conductivity, corresponds to σ, and dT/dx, the temperature gradient, corresponds to dV/dx, the potential gradient. For pure metals there is more than a formal mathematical analogy between Eqs. 29-14 and 29-15. Both heat energy and charge are carried by the free electrons in such metals; empirically, a good electrical conductor (silver, say) is also a good heat conductor, and the electrical conductivity σ is directly related to the thermal conductivity k. ■

Temperature Variation of Resistivity (Optional)

Figure 29-8 shows a summary of some experimental measurements of the resistivity of copper at different temperatures. For practical use of this information, it would be helpful to express it in the form of an equation. Over a limited range of temperature, the relationship between resistivity and temperature is nearly linear. We can fit a straight line to any selected region of Fig. 29-8, using two points to determine the slope of the line. Choosing a reference point, such as that labeled T_0, ρ_0 in the figure, we can express the resistivity ρ at an arbitrary temperature T from the empirical equation of the straight line in Fig. 29-8, which is

$$\rho - \rho_0 = \rho_0 \alpha_{\text{av}} (T - T_0). \qquad (29\text{-}16)$$

(This expression is very similar to that for linear thermal expansion, $\Delta L = \alpha L \, \Delta T$, which we introduced in Section 21-4.) We have written the slope of this line as $\rho_0 \alpha_{\text{av}}$. If we solve Eq. 29-16 for α_{av}, we obtain

$$\alpha_{\text{av}} = \frac{1}{\rho_0} \frac{\rho - \rho_0}{T - T_0}. \qquad (29\text{-}17)$$

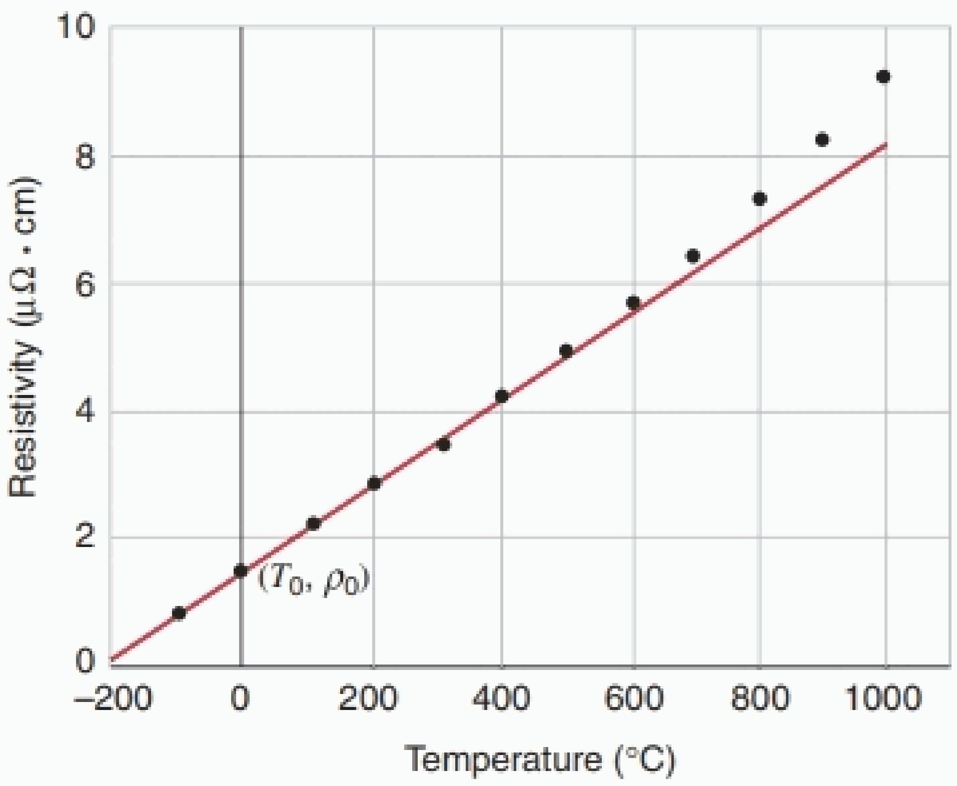


FIGURE 29-8. The dots show selected measurements of the resistivity of copper at different temperatures. Over any given range of temperature, the variation in the resistivity with T can be approximated by a straight line; for example, the line shown fits the data from about -100°C to 400°C.

The quantity α_{av} is the *mean (or average) temperature coefficient of resistivity* over the region of temperature between the two points used to determine the slope of the line. We can define a more general temperature coefficient of resistivity as

$$\alpha = \frac{1}{\rho} \frac{d\rho}{dT}, \qquad (29\text{-}18)$$

which is the fractional change in resistivity $d\rho/\rho$ per change in temperature dT. That is, α gives the dependence of resistivity on temperature *at a particular temperature,* whereas α_{av} gives the average dependence *over a particular interval.* The coefficient α is in general dependent on temperature.

For most practical purposes, Eq. 29-16 gives results that are within the acceptable range of accuracy. Typical values of α_{av} are given in Table 29-1. For more precise work, such as the use of the platinum resistance thermometer to measure temperature (see Section 21-3), the linear approximation is not sufficient. In this case we can add terms in $(T - T_0)^2$ and $(T - T_0)^3$ to the right side of Eq. 29-16 to improve the precision. The coefficients of these additional terms must be determined empirically, in analogy with the coefficient α_{av} in Eq. 29-16. ■

29-5 OHM'S LAW: A MICROSCOPIC VIEW

As we discussed previously, Ohm's law is not a fundamental law of electromagnetism because it depends on the properties of the conducting medium. The law is very simple in form, and it is curious that many materials obey it so well, whereas other materials do not obey it at all. Let us see if we can understand why metals obey Ohm's law—that is, why their resistivities ρ are constant (and not, for example, dependent on the applied electric field).

In a metal, the valence electrons are not attached to individual atoms but are free to move about within the lattice and are called *conduction electrons.* In copper there is one such electron per atom, the other 28 remaining bound to the copper nuclei to form ionic cores.

The theory of electrical conduction in metals is often based on the *free-electron model,* in which (as a first approximation) the conduction electrons are assumed to move throughout the conducting material, somewhat like molecules of gas in a container. In fact, the assembly of conduction electrons is sometimes called an *electron gas.* As we shall see, however, we cannot neglect the effect of the ion cores on this "gas."

The classical Maxwellian velocity distribution (see Section 22-4) for the electron gas would suggest that the conduction electrons have a broad distribution of velocities from zero to infinity, with a well-defined average. However, in considering the electrons we cannot ignore quantum mechanics, which gives a very different view. In the quantum distribution (see Section 49-4), the electrons that readily contribute to electrical conduction are concentrated in a very narrow interval of kinetic energies and therefore of speeds. To a very good approximation, we can assume that the electrons move with a uniform average speed. In the case of copper, this speed is about $v_{\mathrm{av}} = 1.6 \times 10^6$ m/s. Furthermore, whereas the Maxwellian average speed depends strongly on the temperature, the effective speed obtained from the quantum distribution is nearly independent of temperature.

In the absence of an electric field, the electrons move randomly, again like the molecules of gas in a container. Occasionally, an electron collides with an ionic core of the lattice, suffering a sudden change in direction in the process. As we did in the case of collisions of gas molecules, we can associate a mean free path λ and a mean free time τ to the average distance and time between collisions. (Collisions between the electrons themselves are rare and do not affect the electrical properties of the conductor.)

In an ideal metallic crystal (containing no defects or impurities) at 0 K, electron–lattice collisions would not occur, according to the predictions of quantum physics; that is, $\lambda \rightarrow \infty$ as $T \rightarrow 0$ K for ideal crystals. Collisions take place in actual crystals because (1) the ionic cores at any temperature T are vibrating about their equilibrium positions in a random way; (2) impurities—that is, foreign atoms—may be present; and (3) the crystal may contain lattice imperfections, such as missing atoms and displaced atoms. Consequently, the resistivity of a metal can be increased by (1) raising its temperature, (2) adding small amounts of impurities, and (3) straining it severely, as by drawing it through a die, to increase the number of lattice imperfections.

When we apply an electric field to a metal, the electrons modify their random motion in such a way that they drift slowly, in the opposite direction to that of the field, with an average drift speed v_{d}. This drift speed is very much less (by a factor of something like 10^{10}; see Sample Problem 29-3) than the effective average speed v_{av}. Figure 29-9 suggests the relationship between these two speeds. The solid lines suggest a possible random path followed by an electron in the absence of an applied field; the electron proceeds from x to y, making six collisions on the way. The dashed lines show how this same event *might* have occurred if an electric field $\vec{\mathbf{E}}$ had been applied. Note that the electron drifts steadily to the right, ending at y' rather than at y. In preparing Fig. 29-9, it has been assumed that the drift speed v_{d} is $0.02 v_{\mathrm{av}}$; actually, it is more like $10^{-10} v_{\mathrm{av}}$, so that the "drift" exhibited in the figure is greatly exaggerated.

We can calculate the drift speed v_{d} in terms of the applied electric field E and of v_{av} and λ. When a field is applied to an electron in the metal, it experiences a force eE, which imparts to it an acceleration a given by Newton's second law,

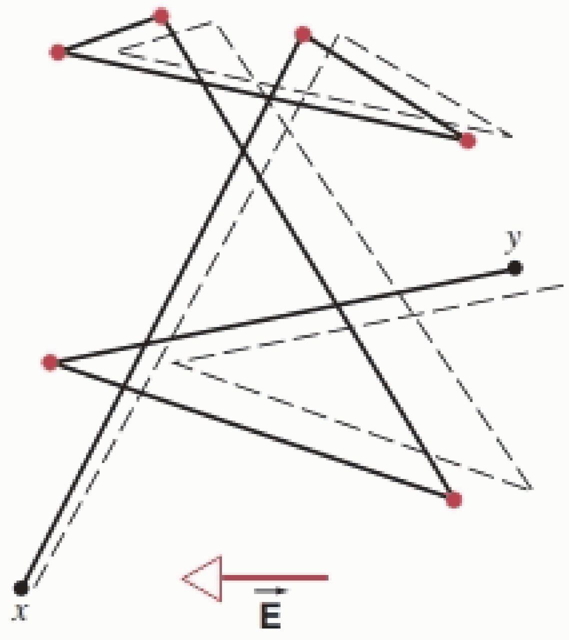


FIGURE 29-9. The solid line segments show an electron moving from x to y, making six collisions en route. The dashed lines show what its path *might* have been in the presence of an applied electric field $\vec{\mathbf{E}}$. Note the gradual but steady drift in the direction of $-\vec{\mathbf{E}}$. (Actually, the dashed lines should be slightly curved to represent the parabolic paths followed by the electrons between collisions.)

$$a = \frac{eE}{m}.$$

Consider an electron that has just collided with an ion core. The collision, in general, momentarily destroys the tendency to drift, and the electron has a truly random direction after the collision. During the time interval to the next collision, the electron's speed changes, on the average, by an amount $a(\lambda / v_{\mathrm{av}})$ or $a\tau$, where τ is the mean time between collisions. We identify this with the drift speed v_{d}, or*

$$v_{\mathrm{d}} = a\tau = \frac{eE\tau}{m}. \tag{29-19}$$

We may also express v_{d} in terms of the current density (Eq. 29-6), which gives

$$v_{\mathrm{d}} = \frac{j}{ne} = \frac{eE\tau}{m}.$$

Combining this with Eq. 29-10 ($\rho = E/j$), we finally obtain

$$\rho = \frac{m}{ne^2\tau}. \tag{29-20}$$

---

*It may be tempting to write Eq. 29-19 as $v_{\mathrm{d}} = \frac{1}{2}a\tau$, reasoning that $a\tau$ is the electron's *final* velocity, and thus that its *average* velocity is half that value. The extra factor of $\frac{1}{2}$ would be correct if we followed a typical electron, taking its drift speed to be the average of its velocity over its mean time τ between collisions. However, the drift speed is proportional to the current density j and must be calculated from the average velocity of *all* the electrons taken at one instant of time. For each electron, the velocity at any time is at, where t is the time since the last collision for that electron. Since the acceleration a is the same for all electrons, the average value of at at a given instant is $a\tau$, where τ is the average time since the last collision, which is the same as the mean time between collisions. For a discussion of this point, see *Electricity and Magnetism,* 2nd ed., by Edward Purcell (McGraw-Hill, 1985), Section 4.4. See also "Drift Speed and Collision Time," by Donald E. Tilley, *American Journal of Physics,* June 1976, p. 597.

Note that m, n, and e in this equation are constants. Thus Eq. 29-20 can be taken as a statement that metals obey Ohm's law if we can show that τ is a constant. In particular, we must show that τ does not depend on the applied electric field E. In this case ρ does not depend on E, which is the criterion that a material obey Ohm's law. The quantity τ depends on the speed distribution of the conduction electrons. We have seen that this distribution is affected only very slightly by the application of even a relatively large electric field, since $v_{\rm av}$ is of the order of 10^6 m/s, and $v_{\rm d}$ (see Sample Problem 29-3) is only of the order of 10^{-4} m/s, a ratio of 10^{10}. Whatever the value of τ is (for copper at 20°C, say) in the absence of a field, it remains essentially unchanged when the field is applied. Thus the right side of Eq. 29-20 is independent of E (which means that ρ is independent of E), and the material obeys Ohm's law.

SAMPLE PROBLEM 29-6. (*a*) What is the mean free time τ between collisions for the conduction electrons in copper? (*b*) What is the mean free path λ for these collisions? Assume an effective speed $v_{\rm av}$ of 1.6×10^6 m/s.

Solution (*a*) From Eq. 29-20 we have

$$\begin{aligned}\tau &= \frac{m}{ne^2\rho}\\ &= \frac{9.11 \times 10^{-31}\ \text{kg}}{(8.49 \times 10^{28}\ \text{m}^{-3})(1.60 \times 10^{-19}\ \text{C})^2(1.69 \times 10^{-8}\ \Omega\cdot\text{m})}\\ &= 2.48 \times 10^{-14}\ \text{s}.\end{aligned}$$

The value of n, the number of conduction electrons per unit volume in copper, was obtained from Sample Problem 29-3; the value of ρ comes from Table 29-1.

(*b*) We define the mean free path from

$$\begin{aligned}\lambda = \tau v_{\rm av} &= (2.48 \times 10^{-14}\ \text{s})(1.6 \times 10^6\ \text{m/s})\\ &= 4.0 \times 10^{-8}\ \text{m} = 40\ \text{nm}.\end{aligned}$$

This is about 150 times the distance between nearest-neighbor ions in a copper lattice. A full treatment based on quantum physics reveals that we cannot view a "collision" as a direct interaction between an electron and an ion. Rather, it is an interaction between an electron and the thermal vibrations of the lattice, lattice imperfections, or lattice impurity atoms. An electron can pass very freely through an "ideal" lattice—that is, a geometrically "perfect" lattice close to the absolute zero of temperature. Mean free paths as large as 10 cm have been observed under such conditions.

29-6 AN INSULATOR IN AN ELECTRIC FIELD

So far we have been talking only about the behavior of conducting materials in electric fields. We now consider what happens when we apply an external electric field to an insulating material. That is, we shall repeat the experiment of Fig. 29-1 with the conducting material replaced by an insulating material.

In an insulator, the electric charges are not free to move. No current results when an insulator is placed in an electric field. The electrons remain firmly locked to their atoms or molecules. Instead of moving charges through the material, all the electric field can do in an insulator is to produce a slight rearrangement of the electric charges within the atoms. However, this small effect can have a substantial influence on the electric field in an insulator.

We begin by considering an insulator such as pure water. The water molecule has a permanent electric dipole moment, as we illustrated in Fig. 26-20. When a water molecule, with its electric dipole moment, is placed in an electric field, as in Fig. 26-19, the field exerts a torque on the dipole that tries to align it with the field. Figure 29-10 shows a collection of dipoles, which have been rotated into alignment by an external field.

To an external observer, the collection of dipoles in Fig. 29-10*b* appears to show negative charges on its upper surface and positive charges on its lower surface. In this respect the insulator resembles the conductor of Fig. 29-1, but the explanation is very different—there is no movement of electrons through the insulating material. In an insulator, an external electric field causes charges to move only over distances that are less than an atomic diameter.

Figure 29-11*a* shows a slab of insulating material that has been placed in an externally applied electric field $\vec{\mathbf{E}}_0$. As a result of the rotation of the dipole moments, there is an apparent sheet of positive charge on the lower surface of the material and a sheet of negative charge on the upper surface. These two sheets of *induced surface charge* establish an electric field $\vec{\mathbf{E}}'$ in the insulator that opposes the applied field, as shown in Fig. 29-11*b*. The effect of aligning the dipoles in the insulator is called *polarization,* and the field $\vec{\mathbf{E}}'$ is known as the polarization field.

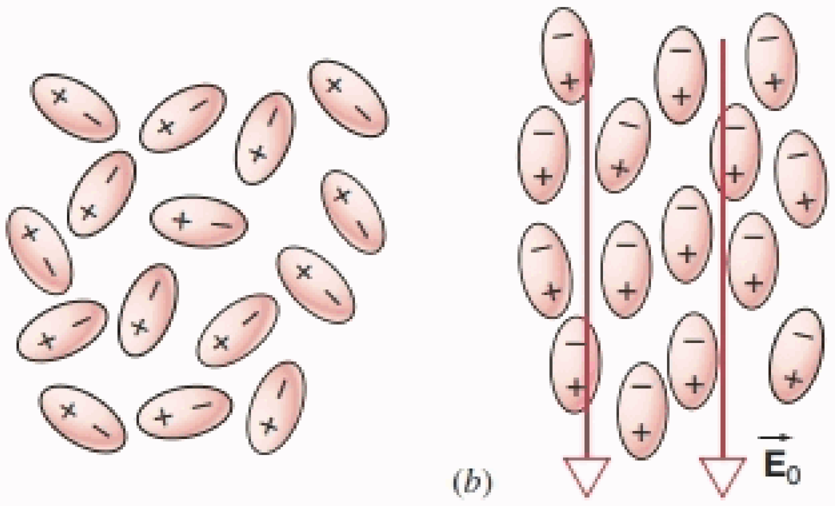
The net field $\vec{\mathbf{E}}$ inside the insulator is the vector sum of the applied field $\vec{\mathbf{E}}_0$ and the polarization field $\vec{\mathbf{E}}'$:

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 + \vec{\mathbf{E}}'. \tag{29-21}$$

Because $\vec{\mathbf{E}}_0$ and $\vec{\mathbf{E}}'$ are in opposite directions, we can write the vector sum as a difference of magnitudes:

$$E = E_0 - E'. \tag{29-22}$$

Figure 29-11*c* shows the net field inside the insulator, which is smaller than the applied field. *When an insulator is placed in an electric field, induced surface charges appear that tend to weaken the original field within the material.*



(*a*) (*b*)

FIGURE 29-10. (*a*) A collection of randomly oriented dipoles. (*b*) An external electric field aligns the dipoles.

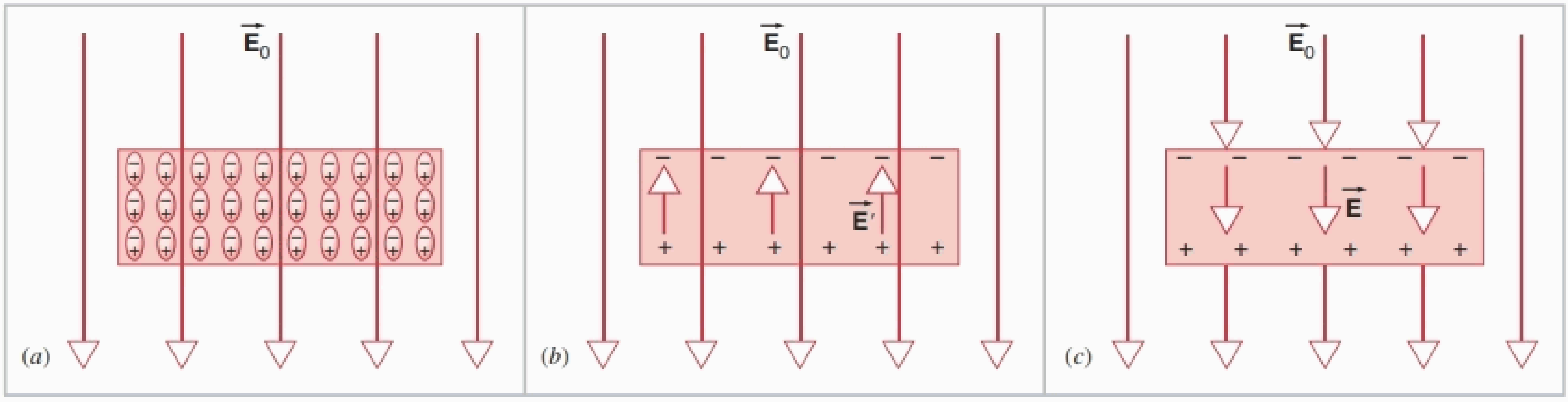


FIGURE 29-11. (*a*) When an insulator is placed in an external field, the dipoles become aligned. (*b*) Induced surface charges on the insulator establish a polarization field $\vec{E}'$ in its interior. (*c*) The net field $\vec{E}$ in the insulator is the vector sum of $\vec{E}_0$ and $\vec{E}'$.

As we increase the applied field $\vec{E}_0$, the polarization field will generally increase. The dipoles in the insulator are in random thermal motion, which tends to destroy their alignment. The greater the applied field, the greater is the torque on the dipoles, the greater is their degree of alignment, and the greater is the polarization field. For many materials, which are called *linear* materials, the polarization field increases in direct proportion to the applied field: $E' \propto E_0$. Using Eq. 29-22, we can also write this proportionality as $E \propto E_0$ and, introducing a constant of proportionality, we have

$$E = \frac{1}{\kappa_e} E_0, \tag{29-23}$$

where the dimensionless constant κ_e is called the *dielectric constant* of the material. The dielectric constant is greater than 1, and so the net field E in the insulator is smaller than the applied field. Like the conductivity or the resistivity, the dielectric constant is characteristic of the type of material (and its temperature), and is independent of the size or shape of any particular object made from the material.

Insulating materials are also known as *dielectric materials,* and we shall use the two terms interchangeably. Table 29-2 shows values of the dielectric constants for various materials at room temperature. Materials with large dielectric constants have large polarization fields, and therefore the fields in their interiors are considerably reduced compared with the applied field.

If we apply a large enough electric field to an insulating material, we can ionize atoms or molecules of the insulator and thus create a condition for electric charge to flow, as in a conductor. The fields necessary for the *breakdown* of various insulators, called the *dielectric strengths,* are given in Table 29-2.

Water is an example of a *polar dielectric* material, because its molecules have permanent electric dipole moments. Effects similar to those described in this section will also occur for *nonpolar* dielectrics, whose molecules do not have permanent dipole moments. Figure 29-12 shows the effect of an electric field on an atom. The atom can be considered as a spherically symmetric cloud of negative charge

TABLE 29-2 **Some Properties of Dielectrics**[a]

Material	*Dielectric Constant* κ_e	*Dielectric Strength* (kV/mm)
Vacuum	1 (exact)	∞
Air (1 atm)	1.00059	3
Polystyrene	2.6	24
Paper	3.5	16
Transformer oil	4.5	12
Pyrex	4.7	14
Mica	5.4	160
Porcelain	6.5	4
Silicon	12	
Water (25°C)	78.5	
Water (20°C)	80.4	
Titania ceramic	130	
Strontium titanate	310	8

[a] Measured at room temperature.

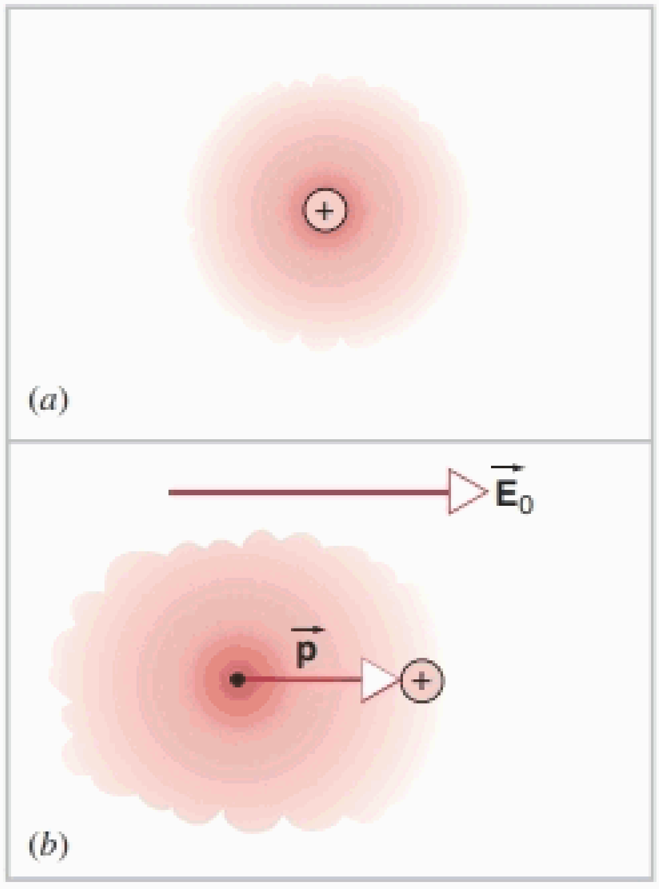


FIGURE 29-12. (*a*) An atom is represented by its positively charged nucleus and its diffuse, negatively charged electron cloud. The centers of positive and negative charge coincide. (*b*) When the atom is placed in an external electric field, the positive and negative charges experience forces in opposite directions, and the centers of the positive and negative charges no longer coincide. The atom acquires an induced dipole moment.

(the electrons) surrounding the positively charged nucleus. With no applied field, the centers of the distributions of positive and negative charge coincide, and the atom has no dipole moment. The electric field causes a separation of charge, as the electrons experience a force in one direction due to the field and the nucleus experiences a force in the opposite direction. The atom acquires an *induced dipole moment* as a result of the action of the electric field. The induced dipole moment disappears when the electric field is removed. The magnitude of this dipole moment is proportional to the applied field, and when the effect of all of the induced dipoles in the material is taken into account, we have again a polarization field E' that is proportional to the applied field, for ordinary field strengths. The induced dipole moment is often responsible for the attraction of a charged object for an uncharged insulator, such as the charged comb and the bits of paper shown in Fig. 25-5.

Because all expressions for electric fields in empty space due to various charge distributions include a factor of $1/\epsilon_0$, Eq. 29-23 suggests that expressions for electric fields in matter will include the factor $1/\kappa_e\epsilon_0$. Since this factor occurs frequently, it is designated by the symbol ϵ:

$$\epsilon = \kappa_e\epsilon_0. \tag{29-24}$$

ϵ is called the *permittivity* of the material (recall that the electric constant ϵ_0 is also known as the permittivity of free space). We can often change equations for electric fields in empty space to apply to electric fields in matter by replacing ϵ_0 with ϵ.

SAMPLE PROBLEM 29-7. Two circular conducting plates of radius 4.2 cm and separation 0.65 cm carry a uniformly distributed charge density of magnitude 2.88×10^{-7} C/m^2, with one plate positively charged and the other negatively charged. The space between the plates is filled with a pyrex glass disk, whose dielectric constant (see Table 29-2) is 4.7. (*a*) Find the electric field in the glass. (*b*) Find the induced charge density on the surfaces of the glass. Consider locations near the center of the disks, where the fields are uniform.

Solution (*a*) In the absence of the dielectric, the electric field due to each circular plate would be $\sigma/2\epsilon_0$, as given by Eq. 26-20. The fields due to the two plates are in the same direction, so they add to give a net field of

$$E_0 = \frac{\sigma}{\epsilon_0} = \frac{2.88 \times 10^{-7}\ \mathrm{C/m^2}}{8.85 \times 10^{-12}\ \mathrm{C^2/N\cdot m^2}} = 3.25 \times 10^4\ \mathrm{N/C}.$$

With the dielectric present, the net field is

$$E = \frac{E_0}{\kappa_e} = \frac{3.25 \times 10^4\ \mathrm{N/C}}{4.7} = 6.9 \times 10^3\ \mathrm{N/C}.$$

(*b*) The polarization field due to the induced surface charge is

$$\begin{aligned} E' = E_0 - E &= 3.25 \times 10^4\ \mathrm{N/C} - 6.9 \times 10^3\ \mathrm{N/C} \\ &= 2.56 \times 10^4\ \mathrm{N/C}. \end{aligned}$$

The two sheets of induced charges set up the electric field E' just as the two sheets of free charges set up the field E_0. With $E' = \sigma_{\text{ind}}/\epsilon_0$, we have

$$\begin{aligned} \sigma_{\text{ind}} = \epsilon_0 E' &= (8.85 \times 10^{-12}\ \mathrm{C^2/N\cdot m^2})(2.56 \times 10^4\ \mathrm{N/C}) \\ &= 2.27 \times 10^{-7}\ \mathrm{C/m^2}. \end{aligned}$$

MULTIPLE CHOICE

29-1 Types of Materials

29-2 A Conductor in an Electric Field: Static Conditions

1. A triangular conductor is placed in an originally uniform electric field. (*a*) Which picture in Fig. 29-13 best represents the static electric field lines near the conductor? (*b*) Which picture in Fig. 29-13 best represents the static equipotential lines near the conductor?

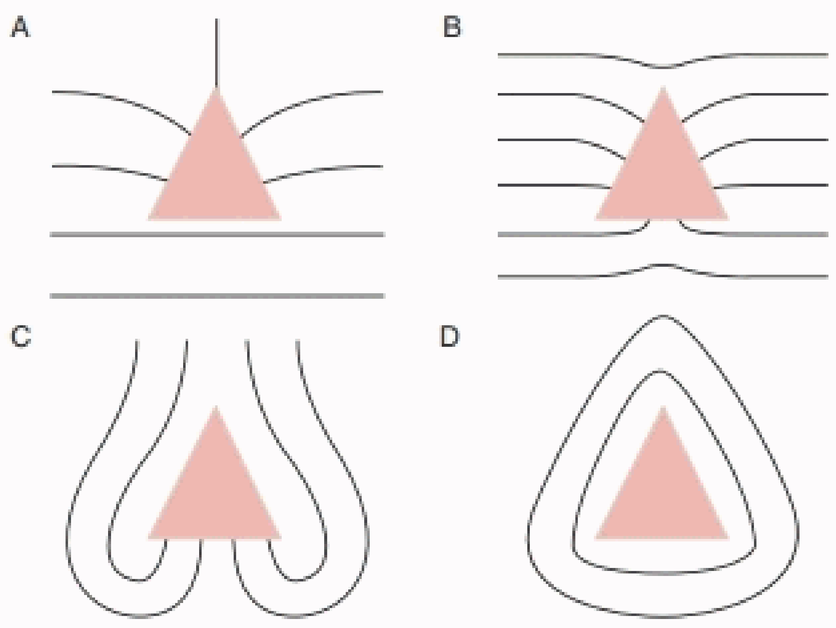


FIGURE 29-13. Multiple-choice question 1.

2. A point charge is placed inside an uncharged spherical conducting shell. Which picture in Fig. 29-14 best shows the electric field lines?

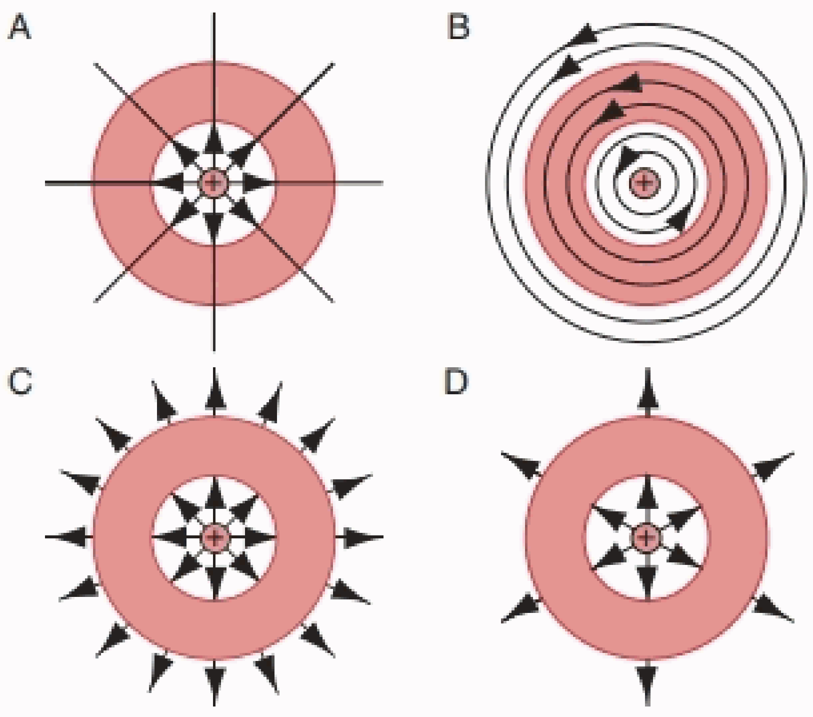


FIGURE 29-14. Multiple-choice question 2.

3. A point charge is placed inside an uncharged spherical conducting shell. Which picture in Fig. 29-15 best shows the electric field lines?

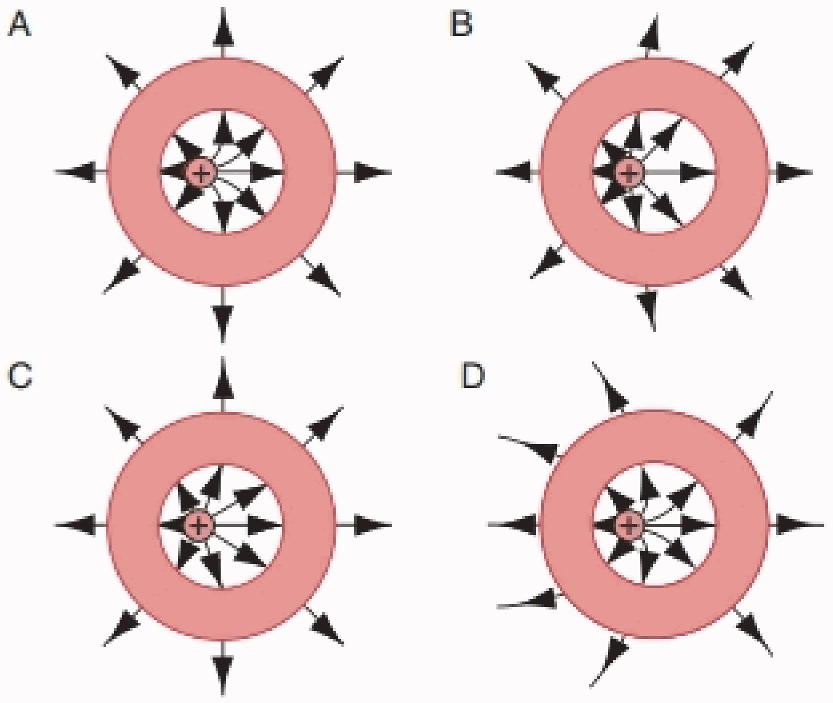


FIGURE 29-15. Multiple-choice question 3.

29-3 A Conductor in an Electric Field: Dynamic Conditions

4. Both current and current density have directions associated with them. Are they vectors?

(A) Only current is a vector.
(B) Only current density is a vector.
(C) Both current and current density are vectors.
(D) Neither current nor current density is a vector.

5. A constant current flows through a conical conductor as shown in Fig. 29-16. End surfaces S_1 and S_2 are two different equipotential surfaces.

(*a*) Through which plane does the greatest current flow?

(A) 1 (B) 2 (C) 3 (D) 4
(E) The current is the same through all.

(*b*) Through which plane is the greatest electric flux?

(A) 1 (B) 2 (C) 3 (D) 4
(E) The electric flux is the same through all.

(*c*) How does the magnitude of the electric field E vary along the central axis moving from S_1 to S_2?

(A) E is constant. (B) E increases.
(C) E decreases.

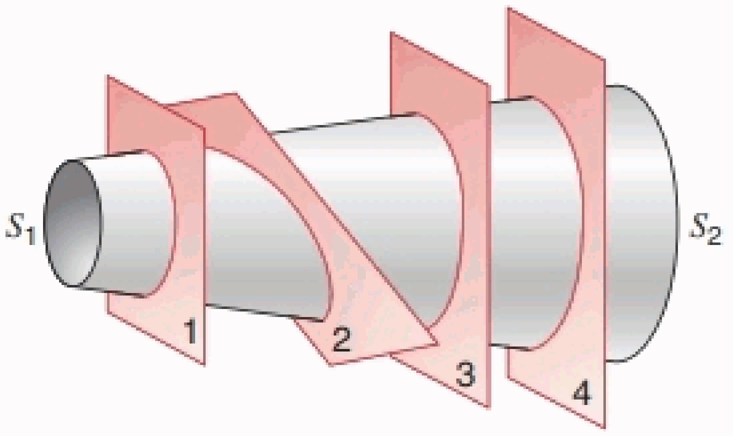


FIGURE 29-16. Multiple-choice question 5.

6. A current flows through of a long cylindrical conductor. In which direction does the current flow?

(A) Toward the end with the higher potential
(B) Toward the end with the lower potential
(C) Neither (A) nor (B), since the surface of a conductor is an equipotential

29-4 Ohmic Materials

7. Two identically shaped wires, A and B, carry identical currents. The wires are made of different substances having differing electron densities, with $n_A > n_B$.

(*a*) Which wire will have the largest current density?

(A) A (B) B (C) The wires are the same.

(*b*) Which wire will have the larger drift speed for the electrons?

(A) A (B) B (C) The wires are the same.

(*c*) Which wire will have the larger electric field E in its interior?

(A) A (B) B (C) The wires are the same.

8. The current–voltage relationship for a certain substance is shown in Fig. 29-17. This substance is ohmic for

(A) all values of ΔV. (B) ΔV between 0 and 3 V.
(C) ΔV greater than 3 V. (D) no values of ΔV.

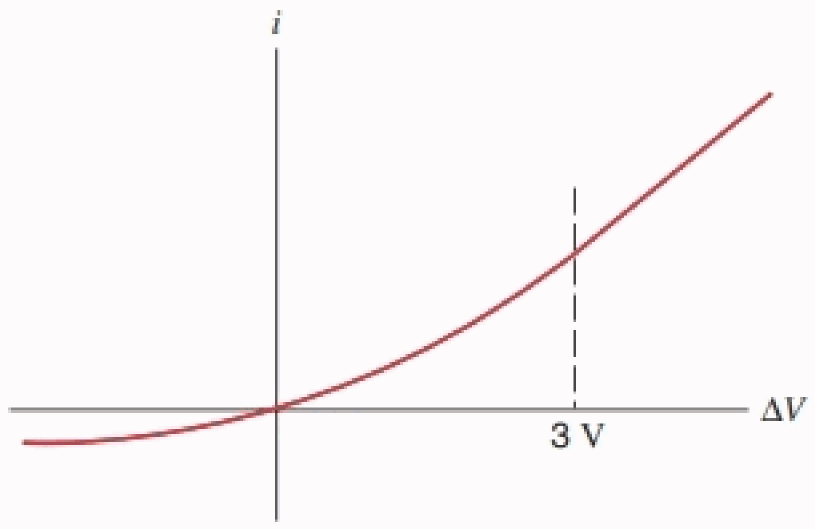


FIGURE 29-17. Multiple-choice question 8.

9. How does the resistance R of an ohmic substance depend on the magnitude E of the applied electric field?

(A) $R \propto E$ (B) ER = a constant
(C) $E + R$ = a constant (D) R is independent of E.

10. A steady current i_{in} flows through the wire that goes into a resistor. A steady current i_{out} flows through the wire that comes out the other end of the resistor.

(*a*) How do i_{in} and i_{out} compare?

(A) $i_{\text{in}} > i_{\text{out}}$ (B) $i_{\text{in}} < i_{\text{out}}$
(C) $i_{\text{in}} = i_{\text{out}}$ always (D) $i_{\text{in}} = i_{\text{out}}$ only if $R = 0$

(*b*) What can be concluded about the potential V_{in} on the end of the wire where the current enters and the potential V_{out} on the end of the wire where the current exits?

(A) $V_{\text{in}} > V_{\text{out}}$ (B) $V_{\text{in}} < V_{\text{out}}$
(C) $V_{\text{in}} = V_{\text{out}}$ always
(D) Nothing, unless more information is given

29-5 Ohm's Law: A Microscopic View

11. How does the drift speed of electrons change as they move through a resistor?

(A) It increases. (B) It decreases.
(C) It remains the same.

12. The resistivity of most conductors increases with temperature. A plausible reason is that, in a conductor,

(A) the electron density changes with temperature.
(B) the charge on each electron changes with temperature.
(C) the time between collisions changes with temperature.
(D) the mass of the electron changes with temperature.

29-6 An Insulator in an Electric Field

13. A spherical insulator is placed in an originally uniform electric field.

(*a*) Which picture in Fig. 29-18 best represents the static electric field lines near and inside the insulator?

(*b*) Which picture in Fig. 29-18 best represents the static equipotential lines near and inside the insulator?

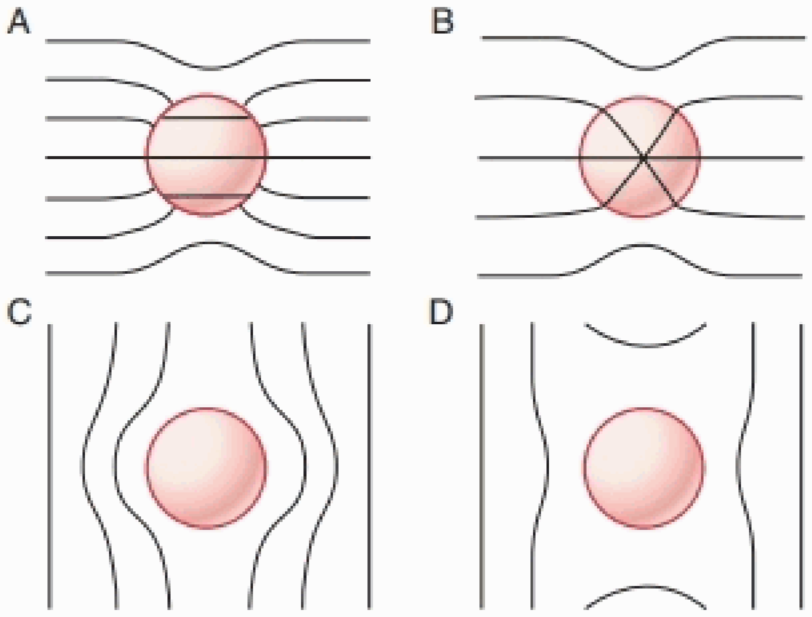


FIGURE 29-18. Multiple-choice question 13.

QUESTIONS

1. Name other physical quantities that, like current, are scalars having a sense represented by an arrow in a diagram.

2. In our convention for the direction of current arrows (*a*) would it have been more convenient, or even possible, to have assumed all charge carriers to be negative? (*b*) Would it have been more convenient, or even possible, to have labeled the electron as positive, the proton as negative, and so on?

3. What experimental evidence can you give to show that the electric charges in current electricity and those in electrostatics are identical?

4. Explain in your own words why we can have $\vec{\mathbf{E}} \neq 0$ inside a conductor in this chapter, whereas we took $\vec{\mathbf{E}} = 0$ for granted in Section 27-6.

5. A current i enters one corner of a square sheet of copper and leaves at the opposite corner. Sketch arrows at various points within the square to represent the relative values of the current density $\vec{\mathbf{j}}$. Intuitive guesses rather than detailed mathematical analyses are called for.

6. Can you see any logic behind the assignment of gauge numbers to household wire? See Exercise 6. If not, then why is this system used?

7. A potential difference ΔV is applied to a copper wire of diameter d and length L. What is the effect on the electron drift speed of (*a*) doubling ΔV, (*b*) doubling L, and (*c*) doubling d?

8. Why is it not possible to measure the drift speed for electrons by timing their travel along a conductor?

9. Describe briefly some possible designs of variable resistors.

10. A potential difference ΔV is applied to a circular cylinder of carbon by clamping it between circular copper electrodes, as in Fig. 29-19. Discuss the difficulty of calculating the resistance of the carbon cylinder using the relation $R = \rho L/A$.

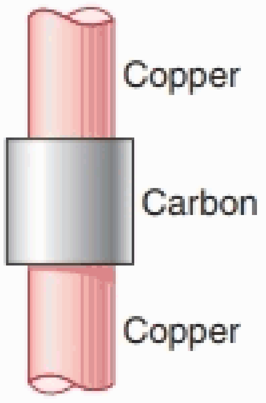


FIGURE 29-19. Question 10.

11. You are given a cube of aluminum and access to two battery terminals. How would you connect the terminals to the cube to ensure (*a*) a maximum and (*b*) a minimum resistance?

12. How would you measure the resistance of a pretzel-shaped block of metal? Give specific details to clarify the concept.

13. Sliding across the seat of an automobile can generate potentials of several thousand volts. Why is the slider not electrocuted?

14. Discuss the difficulties of testing whether the filament of a lightbulb obeys Ohm's law.

15. Will the drift velocity of electrons in a current-carrying metal conductor change when the temperature of the conductor is increased? Explain.

16. Explain why the momentum that conduction electrons transfer to the ions in a metal conductor does not give rise to a resultant force on the conductor.

17. List in tabular form similarities and differences between the flow of charge along a conductor, the flow of water through a horizontal pipe, and the conduction of heat through a slab. Consider such ideas as what causes the flow, what opposes it, what particles (if any) participate, and the units in which the flow may be measured.

18. How does the relation $\Delta V = iR$ apply to resistors that do not obey Ohm's law?

19. A cow and a man are standing in a meadow when lightning strikes the ground nearby. Why is the cow more likely to be killed than the man? The responsible phenomenon is called "step voltage."

20. The lines in Fig. 29-9 should be curved slightly. Why?

21. A fuse in an electrical circuit is a wire that is designed to melt, and thereby open the circuit, if the current exceeds a predetermined value. What are some characteristics of ideal fuse wire?

22. Why does an incandescent lightbulb grow dimmer with use?

23. The character and quality of our daily lives are influenced greatly by devices that do not obey Ohm's law. What can you say in support of this claim?

24. From a student's paper: "The relationship $R = \Delta V/i$ tells us that the resistance of a conductor is directly proportional to the potential difference applied to it." What do you think of this proposition?

25. Carbon has a negative temperature coefficient of resistivity, which means that its resistivity drops as its temperature increases. Would its resistivity disappear entirely at some high enough temperature?

26. Can a dielectric conduct electricity? Can a conductor have dielectric properties?

27. Would you expect the dielectric constant of a material to vary with temperature? If so, how? Does whether or not the molecules have permanent dipole moments matter here?

28. Show that the dielectric constant of a conductor can be taken to be infinitely great.

29. An electric field can polarize gases in several ways: by distorting the electron clouds of molecules, by orienting polar molecules, by bending or stretching the bonds in polar molecules. How does this differ from polarization of molecules in liquids and solids?

30. A dielectric object in a nonuniform electric field experiences a net force. Why is there no net force if the field is uniform?

31. A stream of tap water can be deflected if a charged rod is brought close to the stream. Explain carefully how this happens.

EXERCISES

29-1 Types of Materials

29-2 A Conductor in an Electric Field: Static Conditions

29-3 A Conductor in an Electric Field: Dynamic Conditions

1. A current of 4.82 A exists in a 12.4 Ω resistor for 4.60 min. (*a*) How much charge and (*b*) how many electrons pass through any cross section of the resistor in this time?

2. The current in the electron beam of a typical video display terminal is 200 μA. How many electrons strike the screen each minute?

3. Suppose that we have 2.10×10^8 doubly charged positive ions per cubic centimeter, all moving north with a speed of 1.40×10^5 m/s. (*a*) Calculate the current density, in magnitude and direction. (*b*) Can you calculate the total current in this ion beam? If not, what additional information is needed?

4. A small but measurable current of 123 pA exists in a copper wire whose diameter is 2.46 mm. Calculate (*a*) the current density and (*b*) the electron drift speed. See Sample Problem 29-3.

5. Suppose that the material composing a fuse (see Question 21) melts once the current density rises to 440 A/cm^2. What diameter of cylindrical wire should be used for the fuse to limit the current to 0.552 A?

6. The (United States) National Electric Code, which sets maximum safe currents for rubber-insulated copper wires of various diameters, is given (in part) below. Plot the safe current density as a function of diameter. Which wire gauge has the maximum safe current density?

Gauge[a]	4	6	8	10	12	14	16	18
Diameter (mils)[b]	204	162	129	102	81	64	51	40
Safe current (A)	70	50	35	25	20	15	6	3

[a] A way of identifying the wire diameter.

[b] 1 mil = 10^{-3} in.

7. A current is established in a gas discharge tube when a sufficiently high potential difference is applied across the two electrodes in the tube. The gas ionizes; electrons move toward the positive terminal and singly charged positive ions toward the negative terminal. What are the magnitude and direction of the current in a hydrogen discharge tube in which 3.1×10^{18} electrons and 1.1×10^{18} protons move past a cross-sectional area of the tube each second?

8. A *pn* junction is formed from two different semiconducting materials in the form of identical cylinders with radius 0.165 mm, as depicted in Fig. 29-20. In one application, 3.50×10^{15} electrons per second flow across the junction from the *n* to the *p* side while 2.25×10^{15} holes per second flow from the *p* to the *n* side. (A hole acts like a particle with charge $+1.6 \times 10^{-19}$ C.) Find (*a*) the total current and (*b*) the current density.

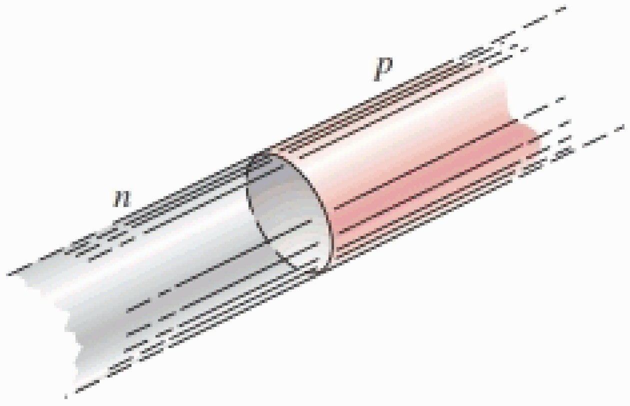


FIGURE 29-20. Exercise 8.

9. Near the Earth, the density of protons in the solar wind is 8.70 cm^{-3} and their speed is 470 km/s. (*a*) Find the current density of these protons. (*b*) If the Earth's magnetic field did not deflect them, the protons would strike the Earth. What total current would the Earth receive?

10. The belt of an electrostatic accelerator is 52.0 cm wide and travels at 28.0 m/s. The belt carries charge into the sphere at a rate corresponding to 95.0 μA. Compute the surface charge density on the belt. See Section 28-10.

11. How long does it take electrons to get from a car battery to the starting motor? Assume that the current is 115 A and the electrons travel through copper wire with cross-sectional area 31.2 mm^2 and length 85.5 cm. See Sample Problem 29-3.

29-4 Ohmic Materials

12. A human being can be electrocuted if a current as small as 50 mA passes near the heart. An electrician working with sweaty hands makes good contact with two conductors being held one in each hand. If the electrician's resistance is 1800 Ω, what might be the fatal voltage difference? (Electricians often work with "live" wires.)

13. A steel trolley-car rail has a cross-sectional area of 56 cm^2. What is the resistance of 11 km of rail? The resistivity of the steel is $3.0 \times 10^{-7}\ \Omega\cdot\text{m}$.

14. From the slope of the line in Fig. 29-8, estimate the average temperature coefficient of resistivity for copper at room temperature and compare with the value given in Table 29-1.

15. A wire 4.0 m long and 6.0 mm in diameter has a resistance of 15 mΩ. A potential difference of 23 V is applied between the ends. (*a*) What is the current in the wire? (*b*) Calculate the current density. (*c*) Calculate the resistivity of the wire material. Can you identify the material? See Table 29-1.

16. The copper windings of a motor have a resistance of 50 Ω at 20°C when the motor is idle. After running for several hours the resistance rises to 58 Ω. What is the temperature of the windings? Ignore changes in the dimensions of the windings. See Table 29-1.

17. Show that if changes in the dimensions of a conductor with changing temperature can be ignored, then the resistance varies with temperature according to $R - R_0 = \alpha_{\text{av}} R_0 (T - T_0)$.

18. A coil is formed by winding 250 turns of insulated gauge 8 copper wire (see Exercise 6) in a single layer on a cylindrical form whose radius is 12.2 cm. Find the resistance of the coil. Neglect the thickness of the insulation. See Table 29-1.

19. Two conductors are made of the same material and have the same length. Conductor *A* is a solid wire of diameter *D*. Conductor *B* is a hollow tube of outside diameter 2*D* and inside diameter *D*. Find the resistance ratio, R_A/R_B, measured between their ends.

20. What must be the diameter of an iron wire if it is to have the same resistance as a copper wire 1.19 mm in diameter, both wires being the same length?

21. An electrical cable consists of 125 strands of fine wire, each having 2.65-μΩ resistance. The same potential difference is applied between the ends of each strand and results in a total current of 750 mA. (*a*) What is the current in each strand? (*b*) What is the applied potential difference? (*c*) What is the resistance of the cable?

22. A copper wire and an iron wire of the same length have the same potential difference applied to them. (*a*) What must be the ratio of their radii if the current is to be the same? (*b*) Can the current density be made the same by suitable choices of the radii?

23. When a potential difference of 115 V is applied between the ends of a 9.66-m-long wire, the current density is 1.42 A/cm^2. Calculate the conductivity of the wire material.

24. In the lower atmosphere of the Earth there are negative and positive ions, created by radioactive elements in the soil and cosmic rays from space. In a certain region, the atmospheric electric field strength is 120 V/m, directed vertically down. Due to this field, singly charged positive ions, 620 per cm^3, drift downward, and singly charged negative ions, 550 per cm^3, drift upward; see Fig. 29-21. The measured conductivity is $2.70 \times 10^{-14}/\Omega\cdot\text{m}$. Calculate (*a*) the ion drift speed, assumed to be the same for positive and negative ions; and (*b*) the current density.

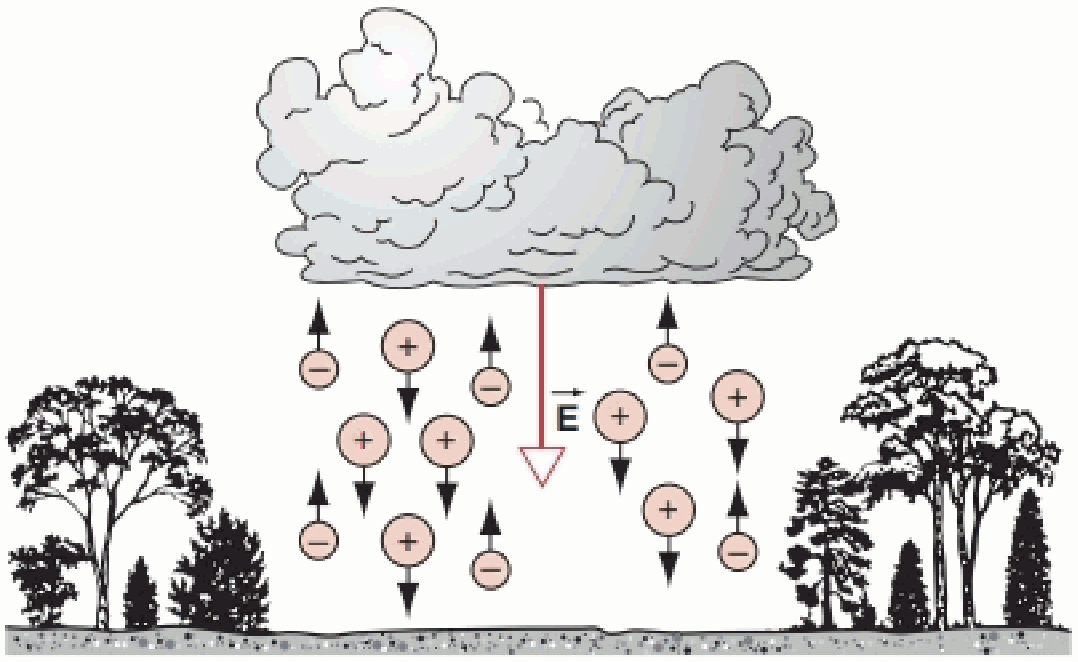


FIGURE 29-21. Exercise 24.

25. Copper and aluminum are being considered for a high-voltage transmission line that must carry a current of 62.3 A. The resistance per unit length is to be 0.152 Ω/km. Compute for each choice of cable material (*a*) the current density and (*b*) the mass of 1.00 m of the cable. The densities of copper and aluminum are 8960 and 2700 kg/m^3, respectively.

26. Using data from Fig. 29-7*b*, plot the resistance of the *pn* junction diode as a function of applied potential difference.

27. For a hypothetical electronic device, the potential difference ΔV, measured across the device, is related to the current *i* by $\Delta V = (3.55 \times 10^6\ \text{V/A}^2)i^2$. (*a*) Find the resistance when the current is 2.40 mA. (*b*) At what value of the current is the resistance equal to 16.0 Ω?

29-5 Ohm's Law: A Microscopic View

28. Calculate the mean free time between collisions for conduction electrons in aluminum at 20°C. Each atom of aluminum contributes three conduction electrons. Take needed data from Table 29-1 and Appendix D. See also Sample Problem 29-3.

29-6 An Insulator in an Electric Field

29. A 1-μC point charge is embedded in the center of a solid Pyrex sphere of radius $R = 10$ cm. (*a*) Calculate the electric field strength *E* just beneath the surface of the sphere. (*b*) Assuming that there are no other *free* charges, calculate the strength of the electric field just outside the surface of the sphere. (*c*) What is the induced surface charge density σ_{ind} on the surface of the Pyrex sphere?

30. Two equal but opposite point charges $+q$ and $-q$ are separated by a distance of 10 cm in air. What value of *q* will provide for an electric field strength midway between the charges that will exceed the dielectric strength of air?

31. A spherical conductor of radius *R* is at a potential *V*; assume $V = 0$ at infinity. (*a*) What is the minimum value of *V* that will result in an electric field strength just above the surface of the sphere that will exceed the dielectric strength of air? (*b*) Is it easier to get a "spark" from a ball at a given potential with a larger or a smaller radius? (*c*) Use your answer to explain why lightning rods are pointed.

PROBLEMS

1. You are given an isolated conducting sphere of 13-cm radius. One wire carries a current of 1.0000020 A into it. Another wire carries a current of 1.0000000 A out of it. How long would it take for the sphere to increase in potential by 980 V?

2. In a hypothetical fusion research lab, high-temperature helium gas is completely ionized, each helium atom being separated into two free electrons and the remaining positively charged nucleus (alpha particle). An applied electric field causes the alpha particles to drift to the east at 25 m/s while the electrons drift to the west at 88 m/s. The alpha particle density is $2.8 \times 10^{15}\ \text{cm}^{-3}$. Calculate the net current density; specify the current direction.

3. A 4.0-cm-long caterpillar crawls in the direction of electron drift along a 5.2-mm-diameter bare copper wire that carries a current of 12 A. (*a*) Find the potential difference between the two ends of the caterpillar. (*b*) Is its tail positive or negative compared to its head? (*c*) How much time could it take the caterpillar to crawl 1.0 cm and still keep up with the drifting electrons in the wire?

4. A steady beam of alpha particles ($q = 2e$) traveling with kinetic energy 22.4 MeV carries a current of 250 nA. (*a*) If the beam is directed perpendicular to a plane surface, how many alpha particles strike the surface in 2.90 s? (*b*) At any instant, how many alpha particles are there in a given 18.0-cm length of the beam? (*c*) Through what potential difference was it necessary to accelerate each alpha particle from rest to bring it to an energy of 22.4 MeV?

5. In the two intersecting storage rings of circumference 950 m at CERN, protons of kinetic energy 28.0 GeV formed beams of current 30.0 A each. (*a*) Find the total charge carried by the protons in each ring. Assume that the protons travel at very nearly the speed of light. (*b*) A beam is deflected out of a ring onto a 43.5-kg copper block. By how much does the temperature of the block rise?

6. (*a*) The current density across a cylindrical conductor of radius R varies according to the equation

$$j = j_0(1 - r/R),$$

where r is the distance from the axis. Thus the current density is a maximum j_0 at the axis $r = 0$ and decreases linearly to zero at the surface $r = R$. Calculate the current in terms of j_0 and the conductor's cross-sectional area $A = \pi R^2$. (*b*) Suppose that, instead, the current density is a maximum j_0 at the surface and decreases linearly to zero at the axis, so that

$$j = j_0 r/R.$$

Calculate the current. Why is the result different from (*a*)?

7. (*a*) At what temperature would the resistance of a copper conductor be double its resistance at 20°C? (Use 20°C as the reference point in Eq. 29-16; compare your answer with Fig. 29-8.) (*b*) Does this same temperature hold for all copper conductors, regardless of shape or size?

8. A common flashlight bulb is rated at 310 mA and 2.90 V, the respective values of the current and voltage under operating conditions. If the resistance of the bulb filament when cold ($T_0 = 20°\text{C}$) is 1.12 Ω, calculate the temperature of the filament when the bulb is on. The filament is made of tungsten. Assume that Eq. 29-16 holds over the temperature range encountered.

9. A wire with a resistance of 6.0 Ω is drawn out through a die so that its new length is three times its original length. Find the resistance of the longer wire, assuming that the resistivity and density of the material are not changed during the drawing process.

10. A block in the shape of a rectangular solid has a cross-sectional area of $3.50\ \text{cm}^2$, a length of 15.8 cm, and a resistance of 935 Ω. The material of which the block is made has 5.33×10^{22} conduction electrons/m^3. A potential difference of 35.8 V is maintained between its ends. (*a*) Find the current in the block. (*b*) Assuming that the current density is uniform, what is its value? Calculate (*c*) the drift velocity of the conduction electrons and (*d*) the electric field in the block.

11. A rod of a certain metal is 1.6 m long and 5.5 mm in diameter. The resistance between its ends (at 20°C) is $1.09 \times 10^{-3}\ \Omega$. A round disk is formed of this same material, 2.14 cm in diameter and 1.35 mm thick. (*a*) What is the material? (*b*) What is the resistance between the opposing round faces, assuming equipotential surfaces?

12. When a metal rod is heated, not only its resistance but also its length and its cross-sectional area change. The relation $R = \rho L/A$ suggests that all three factors should be taken into account in measuring ρ at various temperatures. (*a*) If the temperature changes by 1.0 C°, what fractional changes in R, L, and A occur for a copper conductor? (*b*) What conclusion do you draw? The coefficient of linear expansion is $1.7 \times 10^{-5}/\text{C}°$.

13. It is desired to make a long cylindrical conductor whose temperature coefficient of resistivity at 20°C will be close to zero. If such a conductor is made by assembling alternate disks of iron and carbon, find the ratio of the thickness of a carbon disk to that of an iron disk. (For carbon, $\rho = 3500 \times 10^{-8}\ \Omega\cdot\text{m}$ and $\alpha = -0.50 \times 10^{-3}/\text{C}°$.)

14. A resistor is in the shape of a truncated right circular cone (Fig. 29-22). The end radii are a and b, and the length is L. If the taper is small, we may assume that the current density is uniform across any cross section. (*a*) Calculate the resistance of this object. (*b*) Show that your answer reduces to $\rho L/A$ for the special case of zero taper ($a = b$).

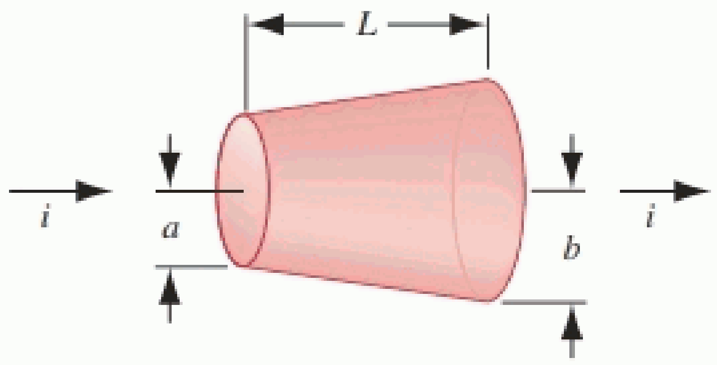


FIGURE 29-22. Problem 14.

15. A resistor is in the shape of a spherical shell, with an inside surface of radius of a covered with a conducting material and an outside surface of radius b covered with a conducting material. Assuming a uniform resistivity ρ, calculate the resistance between the conducting surfaces.

16. Show that, according to the free-electron model of electrical conduction in metals and classical physics, the resistivity of metals should be proportional to $\sqrt{T}$, where T is absolute temperature. (Hint: Treat the electrons as an ideal gas.)

COMPUTER PROBLEM

1. Assume a lightbulb has a tungsten filament that radiates energy at a rate proportional to the temperature difference between the filament and room temperature. Call the constant of proportionality C. Estimate C for a 120-watt bulb in a 120-volt circuit, assuming that all the energy transferred to the filament is radiated as heat transferred to the environment; assume the temperature of the filament is 2500 °C. (*a*) Numerically generate a graph that shows the equilibrium temperature of the bulb as a function of the applied potential difference, remembering that the resistivity of tungsten changes with temperature. (*b*) At what applied voltage must the bulb "burn-out"? (Hint: The filament will melt if it gets too hot.) (*c*) Repeat the above procedure, except now assume that energy is radiated from the bulb according to $k(T^4 - T_0^4)$, where k is a constant you must determine, T_0 is room temperature in kelvin, and T is the filament temperature in kelvin. Compare your results.

CHAPTER 30

CAPACITANCE

In many applications of electric circuits, the goal is to store electrical charge or energy in an electrostatic field. A device that stores charge is called a capacitor, *and the property that determines how much charge it can store is its* capacitance. *We shall see that the capacitance depends on the geometrical properties of the device and not on the electric field or the potential.*

In this chapter we define capacitance and show how to calculate the capacitance of a few simple devices and of combinations of capacitors. We study the energy stored in capacitors and show how it is related to the strength of the electric field. Finally, we investigate how the presence of a dielectric in a capacitor enhances its ability to store electric charge.

30-1 CAPACITORS

A *capacitor** is a device that stores energy in an electrostatic field. A flashbulb, for example, requires a short burst of electric energy that exceeds what a battery can generally provide. A capacitor can draw energy relatively slowly (over several seconds) from the battery, and it then can release the energy rapidly (within milliseconds) through the bulb. Much larger capacitors are used to produce short laser pulses in attempts to induce thermonuclear fusion in tiny pellets of hydrogen. In this case the power level during the pulse is about 10^{14} W, about 200 times the entire electrical generating capacity of the United States, but the pulses typically last only for 10^{-9} s.

Capacitors are also used to produce electric fields, such as the parallel-plate device that gives the very nearly uniform electric field that deflects beams of electrons in a TV or oscilloscope tube.

In circuits, capacitors are often used to smooth out the sudden variations in line voltage that can damage computer memories. In another application, the tuning of a radio or TV receiver is usually done by varying the capacitance of the circuit.

30-2 CAPACITANCE

Figure 30-1 shows a generalized capacitor, consisting of two conductors a and b of arbitrary shape. No matter what their geometry, these conductors are called *plates.* We assume that they are totally isolated from their surroundings. We further assume, for the time being, that the conductors exist in a vacuum.

A capacitor is said to be *charged* if its plates carry equal and opposite charges $+q$ and $-q$. Note that q is *not* the net charge on the capacitor, which is zero. In our discussion of capacitors, we let q represent the absolute value of the charge on either plate; that is, q represents a magnitude only, and the sign of the charge on a given plate must be specified.

We can "charge" a capacitor by connecting one of its plates to the positive terminal of a battery and the other plate to the negative terminal, as shown in Fig. 30-2. As we discuss in the next chapter, the flow of charge in an electrical circuit is analogous to the flow fluid, and the battery serves as a "pump" for electric charge. When we connect a battery to the capacitor (by closing the switch in the

---

*See "Capacitors," by Donald M. Trotter, Jr., *Scientific American,* July 1988, p. 86.

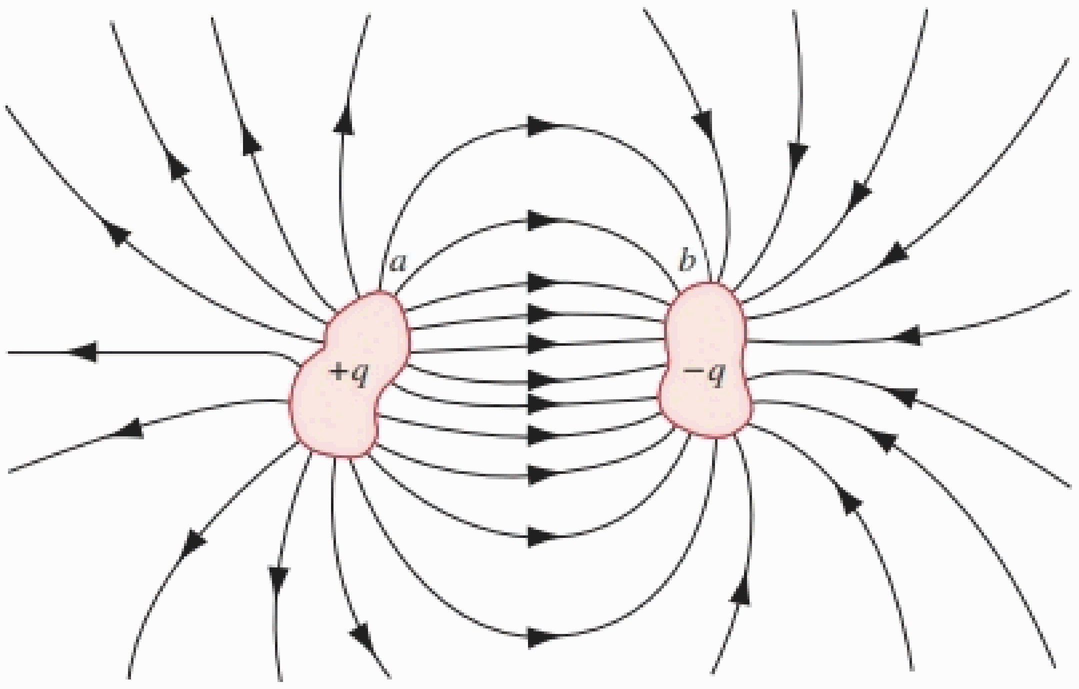


FIGURE 30-1. Two conductors, isolated from one another and from their surroundings, form a capacitor. When the capacitor is charged, the conductors carry equal but opposite charges of magnitude q. The two conductors are called *plates* no matter what their shape.

circuit), the battery "pumps" electrons from the (previously uncharged) positive plate of the capacitor to the negative plate. After the battery moves a quantity of charge of magnitude q, the charge on the positive plate is $+q$ and the charge on the negative plate is $-q$.

An ideal battery maintains a constant potential difference between its terminals. The positive plate and the wire connecting it to the positive terminal of the battery are conductors, and so (under electrostatic conditions) they must be at the same potential V_+ as the positive terminal of the battery. The negative plate and the wire connecting it to the negative terminal of the battery are also conductors, and so (when the switch is closed) they must be at the same potential V_- as the negative terminal of the battery. The potential difference $\Delta V = V_+ - V_-$ between the battery terminals is the same potential difference that appears between the capacitor plates when the switch is closed. We usually describe this as the potential difference "across" the capacitor, meaning the potential difference between its plates.

Figure 30-3 shows the circuit for charging a capacitor by a battery that maintains a constant potential difference $\Delta V = V_+ - V_-$ between its terminals. In a circuit, a capacitor is represented by the symbol ⊣⊢, in which the two parallel lines suggest the two plates of the capacitor.

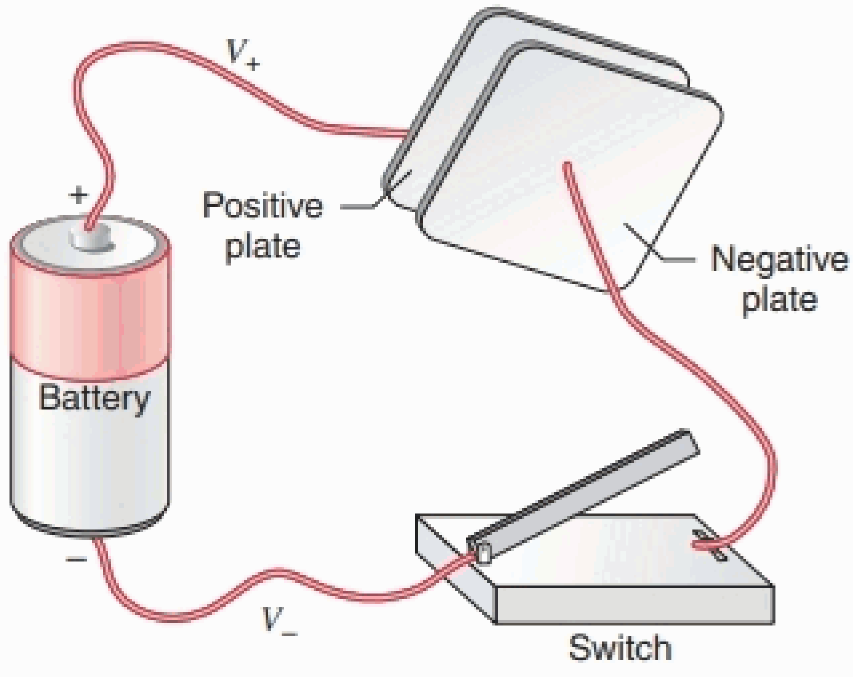


FIGURE 30-2. When the switch is closed, the capacitor becomes charged as the battery moves electrons from the positive plate to the negative plate.

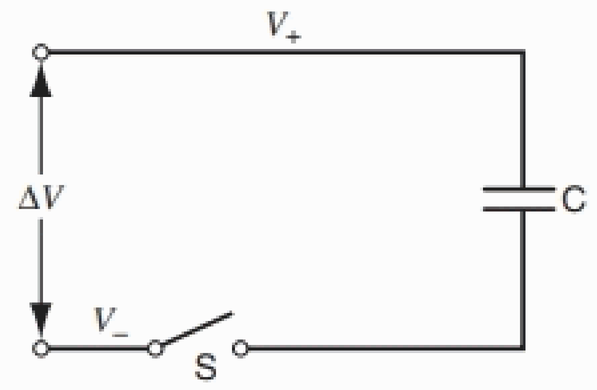


FIGURE 30-3. A schematic circuit diagram equivalent to Fig. 30-2, showing the capacitor C, switch S, and constant potential difference ΔV (supplied by a battery that is not shown in the diagram).

When we charge a capacitor, we find that the charge q that appears on the capacitor plates is always directly proportional to the potential difference ΔV between the plates: $q \propto \Delta V$. The *capacitance C* is the constant of proportionality necessary to make this relationship into an equation:

$$q = C\,\Delta V. \tag{30-1}$$

The capacitance is a geometrical factor that depends on the size, shape, and separation of the plates and on the material that occupies the space between the plates (which for now we assume is a vacuum). The capacitance of a capacitor does *not* depend on ΔV or q.

The SI unit of capacitance that follows from Eq. 30-1 is the coulomb/volt, which is given the name *farad* (abbreviation F):

$$1 \text{ farad} = 1 \text{ coulomb/volt}.$$

The unit is named in honor of Michael Faraday who, among his other contributions, developed the concept of capacitance. The submultiples of the farad, the *microfarad* ($1\ \mu\text{F} = 10^{-6}$ F) and the *picofarad* ($1\text{ pF} = 10^{-12}$ F), are more convenient units in practice. Figure 30-4 shows some capacitors in the microfarad or picofarad range that might be found in electronic or computing equipment.

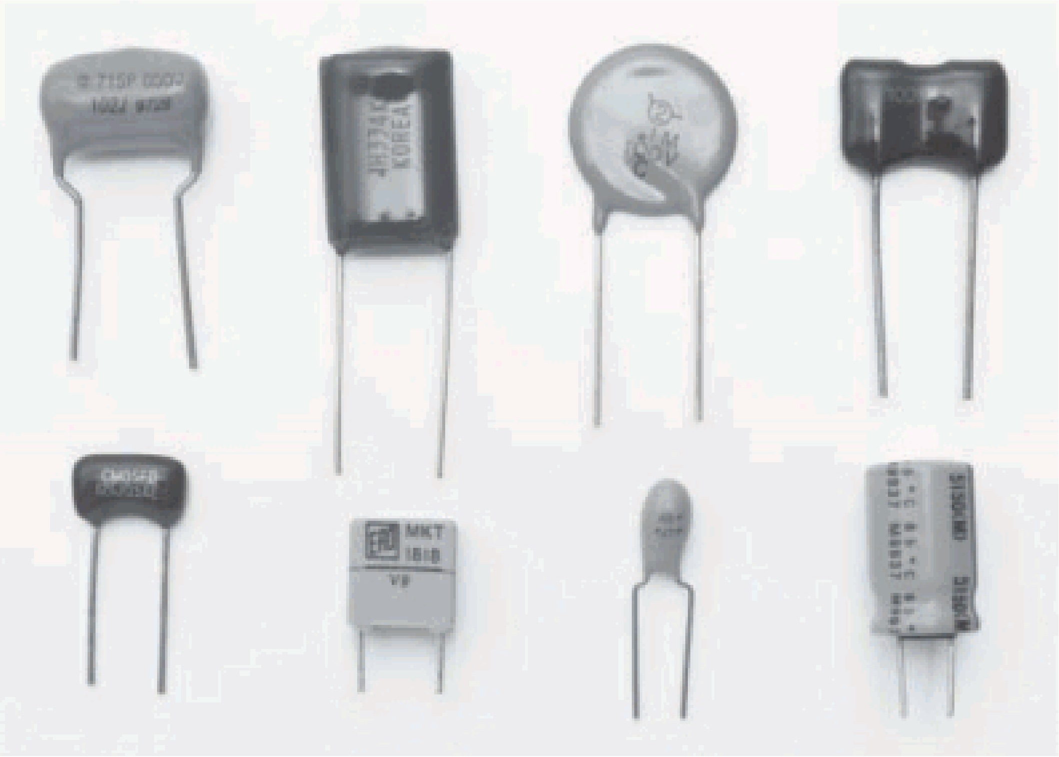


FIGURE 30-4. An assortment of capacitors that might be found in electronic circuits.

SAMPLE PROBLEM 30-1. A storage capacitor on a random access memory (RAM) chip has a capacitance of 0.055 pF. If it is charged to 5.3 V, how many excess electrons are there on its negative plate?

Solution If the negative plate has N excess electrons, it carries a net charge of magnitude $q = Ne$. Using Eq. 30-1, we obtain

$$N = \frac{q}{e} = \frac{C\,\Delta V}{e} = \frac{(0.055 \times 10^{-12}\ \text{F})(5.3\ \text{V})}{1.60 \times 10^{-19}\ \text{C}}$$
$$= 1.8 \times 10^{6}\ \text{electrons}.$$

For electrons, this is a very small number. A speck of household dust, so tiny that it essentially never settles, contains about 10^{17} electrons (and the same number of protons).

Analogy with Fluid Flow (Optional)

In situations involving electric circuits, it is often useful to draw analogies between the movement of electric charge and the movement of material particles such as occurs in fluid flow. In the case of a capacitor, an analogy can be made between a capacitor carrying a charge q and a rigid container of volume v (we use v rather than V for volume so as not to confuse it with potential difference) containing n moles of an ideal gas. The gas pressure p is directly proportional to n for a fixed temperature, according to the ideal gas law (Eq. 21-13)

$$n = \left(\frac{v}{RT}\right)p.$$

For the capacitor (Eq. 30-1)

$$q = C\,\Delta V.$$

Comparison shows that the capacitance C of the capacitor is analogous to the volume v of the container, assuming a fixed temperature for the gas. In fact, the word "capacitor" brings to mind the word "capacity," in the same sense that the volume of a container for gas has a certain "capacity."

We can force more gas into the container by imposing a higher pressure, just as we can force more charge into the capacitor by imposing a higher voltage. Note that any amount of charge can be put on the capacitor, and any mass of gas can be put in the container, up to certain limits. These correspond to electrical breakdown ("arcing over") for the capacitor and to rupture of the walls for the container. ■

30-3 CALCULATING THE CAPACITANCE

Our goal in this section is to calculate the capacitance of a capacitor from its geometry. We do this using the following procedure. (1) We first find the electric field in the region between the plates, using methods such as those described in Section 26-4. (2) We then use Eq. 28-15 to find the potential difference between the positive and negative plates by integrating the electric field along any convenient path connecting the plates:

$$\Delta V = V_+ - V_- = -\int_-^+ \vec{\mathbf{E}}\cdot d\vec{\mathbf{s}} = \int_+^- \vec{\mathbf{E}}\cdot d\vec{\mathbf{s}}. \qquad (30\text{-}2)$$

(3) The outcome of Eq. 30-2 will involve the magnitude of the charge q on the right-hand side. Using Eq. 30-1, we can then find $C = q/\Delta V$.

As we have defined it, ΔV is a positive number. Since q is an absolute magnitude, the capacitance C will always be positive.

We now illustrate this method with several examples.

A Parallel-Plate Capacitor

Figure 30-5 shows a capacitor in which the two flat plates are very large and very close together; that is, the separation d is much smaller than the length or width of the plates. We can neglect the "fringing" of the electric field that occurs near the edges of the plates and assume that the electric field has the same magnitude and direction everywhere in the volume between the plates.

We obtained the electric field for a single large uniformly charged disk at points near its center in Section 26-4: $E = \sigma/2\epsilon_0$. If the capacitor plates are very large, their shape is not important, and we can assume that the electric field due to each plate has this magnitude. The net electric field is the sum of the fields due to the two plates: $\vec{\mathbf{E}} = \vec{\mathbf{E}}_+ + \vec{\mathbf{E}}_-$. As Fig. 30-5 shows, the fields due to the positive and negative plates have the same direction, so we can write

$$E = E_+ + E_- = \sigma/2\epsilon_0 + \sigma/2\epsilon_0 = \sigma/\epsilon_0. \qquad (30\text{-}3)$$

Using $\sigma = q/A$, where A is the surface area of each plate, and substituting Eq. 30-3 into Eq. 30-2, we obtain

$$\Delta V = \int_+^- E\,ds = \frac{q}{\epsilon_0 A}\int_+^- ds = \frac{qd}{\epsilon_0 A}, \qquad (30\text{-}4)$$

where we have chosen an integration path along one of the lines of the electric field, so that $\vec{\mathbf{E}}$ and $d\vec{\mathbf{s}}$ are parallel (see Fig. 30-5).

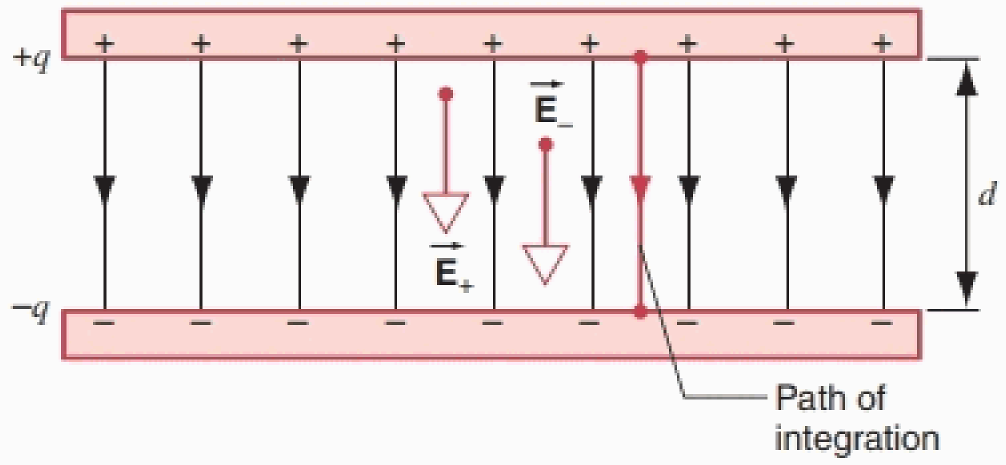


FIGURE 30-5. A parallel-plate capacitor. The path of integration for evaluating Eq. 30-4 is shown.

The capacitance is then obtained from Eq. 30-1: $C = q/\Delta V$, or

$$C = \frac{\epsilon_0 A}{d} \qquad \text{(parallel-plate capacitor).} \qquad (30\text{-}5)$$

You can see from this equation why we say that the capacitance depends on geometrical factors, in this case the plate separation d and area A. The capacitance does not depend on the voltage difference between the plates or the charge carried by the plates.

Note that the right-hand side of Eq. 30-5 has the form of ϵ_0 times a quantity with the dimension of length (A/d). We will find that all expressions for capacitance have essentially this same form, which suggests that the units of ϵ_0 can be expressed as capacitance divided by length:

$$\epsilon_0 = 8.85 \times 10^{-12}\ \text{F/m} = 8.85\ \text{pF/m}.$$

These units for ϵ_0 are often more useful for calculations of capacitance than our previous (and equivalent) units of $\text{C}^2/\text{N} \cdot \text{m}^2$.

A Spherical Capacitor

Figure 30-6 shows a cross section of a spherical capacitor, in which the inner conductor is a solid sphere of radius a, and the outer conductor is a hollow spherical shell of inner radius b. We assume that the inner sphere carries a charge $+q$ and that the outer sphere has a charge $-q$. From our analysis of conductors using Gauss' law (see Section 27-6), we know that the charge on the inner conductor resides on its surface and that the charge on the outer conductor resides on its inner surface. (Draw a spherical Gaussian surface of radius slightly larger than b; the surface lies entirely within the outer conductor, so $E = 0$ everywhere on the surface and the flux through the surface is zero. Therefore the surface encloses no net charge, as Fig. 30-6 shows.)

In the region $a < r < b$, we can use Gauss' law to determine that, in the region between the conductors, the electric field depends only on the charge on the inner sphere, and that this field is the same as that of a point charge at its center (recall the shell theorems discussed in Section 27-5). We therefore have

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} \qquad a < r < b. \qquad (30\text{-}6)$$

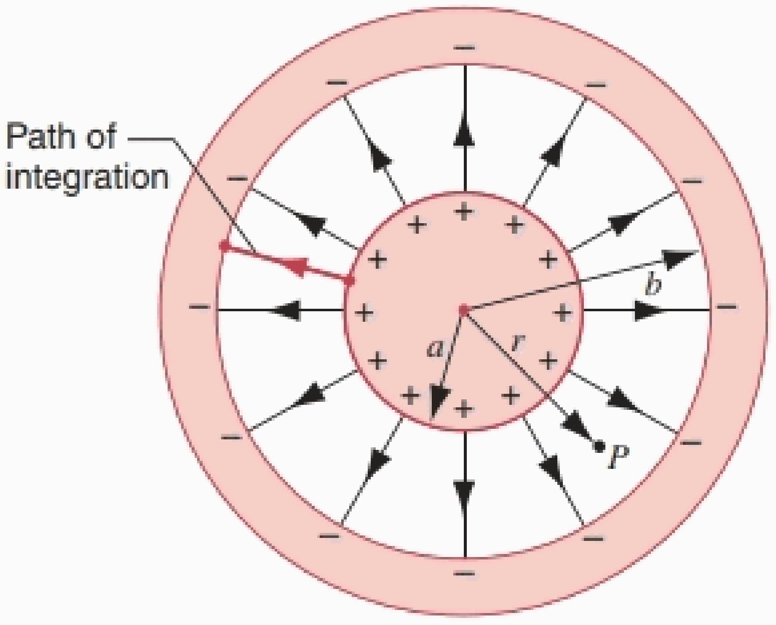


FIGURE 30-6. A cross section through a spherical or a cylindrical capacitor. The electric field at any point P in the interior is due only to the inner conductor. The path of integration for evaluating Eq. 30-7 or Eq. 30-10 is shown.

Substituting this expression for the electric field into Eq. 30-2 and integrating along the path shown in Fig. 30-6 from the positive plate to the negative plate, we obtain

$$\Delta V = \int_{+}^{-} E\,ds = \int_a^b \frac{q}{4\pi\epsilon_0}\frac{dr}{r^2} = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{a} - \frac{1}{b}\right)$$
$$= \frac{q}{4\pi\epsilon_0}\,\frac{b-a}{ab}. \qquad (30\text{-}7)$$

Because the path of integration is in the radial direction, we have $\vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = E\,ds$ and $ds = dr$.

Using $C = q/\Delta V$, we now find

$$C = 4\pi\epsilon_0 \frac{ab}{b-a} \qquad \text{(spherical capacitor).} \qquad (30\text{-}8)$$

Note that the capacitance again has the form of ϵ_0 times a quantity with the dimension of length.

A Cylindrical Capacitor

Figure 30-6 can also represent the cross section of a cylindrical capacitor, in which the inner conductor is a solid rod of radius a carrying a charge $+q$ uniformly distributed over its surface, and the outer conductor is a coaxial cylindrical shell of inner radius b carrying a charge of $-q$ uniformly distributed over its inner surface. The capacitor has length L, and we assume $L \gg b$ so that, as was the case with the parallel-plate capacitor, we can neglect the "fringing" field at the ends of the capacitor.

Just as we used Gauss' law in the spherical geometry to obtain the two shell theorems, we can obtain two similar results in the cylindrical geometry. If only the uniformly charged outer cylindrical conductor were present, we could construct a Gaussian surface in the shape of a long cylinder of radius $r < b$ having the same axis as the outer cylinder. This surface encloses no net charge, so we conclude that $E = 0$ everywhere on the Gaussian surface. As in the case of the spherical shell, a uniformly charged cylindrical shell produces no electric field in its interior. Using a cylindrical Gaussian surface with $r > a$, we can deduce that the inner cylinder behaves just like a uniform line of charge, for which the field points radially outward from the axis and has a magnitude that we calculated in Section 26-4 (Eq. 26-17):

$$E = \frac{1}{2\pi\epsilon_0}\frac{q}{Lr} \qquad a < r < b, \qquad (30\text{-}9)$$

where we have replaced the linear charge density λ with q/L and the distance y with the radial coordinate r. Equation 30-2 now gives

$$\Delta V = \int_{+}^{-} E\,ds = \frac{q}{2\pi\epsilon_0 L}\int_a^b \frac{dr}{r}$$

$$= \frac{q}{2\pi\epsilon_0 L}\ln\left(\frac{b}{a}\right). \quad (30\text{-}10)$$

As we did for the spherical capacitor, we have chosen an integrating path from the positive plate to the negative plate in the radial direction, so $\vec{\mathbf{E}}\cdot d\vec{\mathbf{s}} = E\,ds$ and $ds = dr$.

Equation 30-1 now gives the capacitance:

$$C = 2\pi\epsilon_0 \frac{L}{\ln(b/a)} \quad \text{(cylindrical capacitor)}. \quad (30\text{-}11)$$

Note once again that only geometrical factors appear in this equation and that the capacitance has the form of ϵ_0 multiplied by a quantity with the dimension of length.

SAMPLE PROBLEM 30-2. The plates of a parallel-plate capacitor are separated by a distance $d = 1.0$ mm. What must be the plate area if the capacitance is to be 1.0 F?

Solution From Eq. 30-5 we have

$$A = \frac{Cd}{\epsilon_0} = \frac{(1.0\text{ F})(1.0\times10^{-3}\text{ m})}{8.85\times10^{-12}\text{ F/m}} = 1.1\times10^{8}\text{ m}^2.$$

This is the area of a square more than 10 km on edge. The farad is indeed a large unit. Modern technology, however, has permitted the construction of 1-F capacitors of very modest size. These "Supercaps" are used as backup voltage sources for computers; they can maintain the computer memory for up to 30 days in case of power failure.

SAMPLE PROBLEM 30-3. The space between the conductors of a long coaxial cable, used to transmit TV signals, has an inner radius $a = 0.15$ mm and an outer radius $b = 2.1$ mm. What is the capacitance per unit length of this cable?

Solution From Eq. 30-11 we have

$$\frac{C}{L} = \frac{2\pi\epsilon_0}{\ln(b/a)} = \frac{(2\pi)(8.85\text{ pF/m})}{\ln(2.1\text{ mm}/0.15\text{ mm})} = 21\text{ pF/m}.$$

SAMPLE PROBLEM 30-4. What is the capacitance of the Earth, viewed as an isolated conducting sphere of radius $R =$ 6370 km?

Solution We can assign a capacitance to a single isolated spherical conductor by assuming that the "missing plate" is a conducting sphere of infinite radius.

If we let $b \to \infty$ in Eq. 30-8 and substitute R for a, we find

$$C = 4\pi\epsilon_0 R \quad \text{(isolated sphere)}. \quad (30\text{-}12)$$

Substituting, we obtain

$$C = (4\pi)(8.85\times10^{-12}\text{ F/m})(6.37\times10^{6}\text{ m})$$
$$= 7.1\times10^{-4}\text{ F} = 710\ \mu\text{F}.$$

A tiny 1-F Supercap has a capacitance that is about 1400 times larger than that of the Earth.

30-4 CAPACITORS IN SERIES AND PARALLEL

In analyzing electric circuits, it is often desirable to know the *equivalent capacitance* of two or more capacitors that are connected in a certain way. By "equivalent capacitance" we mean the capacitance of a single capacitor that can be substituted for the combination with no change in the operation of the rest of the circuit.

Capacitors Connected in Parallel

Figure 30-7*a* shows two capacitors connected *in parallel.* There are three properties that characterize a parallel connection of circuit elements. (1) In traveling from *a* to *b*, we can take any of several (two, in this case) *parallel* paths, each of which goes through *only one* of the parallel elements. (2) When a battery of potential difference ΔV is connected across the combination (that is, one terminal of the battery is connected to point *a* in Fig. 30-7*a* and the other terminal to point *b*), the same potential difference ΔV appears across each element of the parallel connection. The wires and capacitor plates are conductors and therefore equipotentials under electrostatic conditions. The potential at *a* appears on the wires connected to *a* and on the two left-hand capacitor plates; similarly, the potential at *b* appears on all the wires connected to *b* and on the two right-hand capacitor plates. (3) The total charge that is delivered by the battery to the combination is shared among the elements; some charge "pumped" by the battery ends up on C_1 and some on C_2.

With these principles in mind, we can now find the equivalent capacitance C_{eq} that gives the same total capacitance between points *a* and *b*, as indicated in Fig. 30-7*b*. We assume a battery of potential difference ΔV to be connected between points *a* and *b*. For each capacitor, we can write (using Eq. 30-1)

$$q_1 = C_1\Delta V \quad \text{and} \quad q_2 = C_2\Delta V. \quad (30\text{-}13)$$

In writing these equations, we have used the same value of the potential difference across the capacitors, in accordance with the second characteristic of a parallel connection stated previously. The battery extracts charge q from one side of the circuit and moves it to the other side. This charge is shared among the two elements according to the

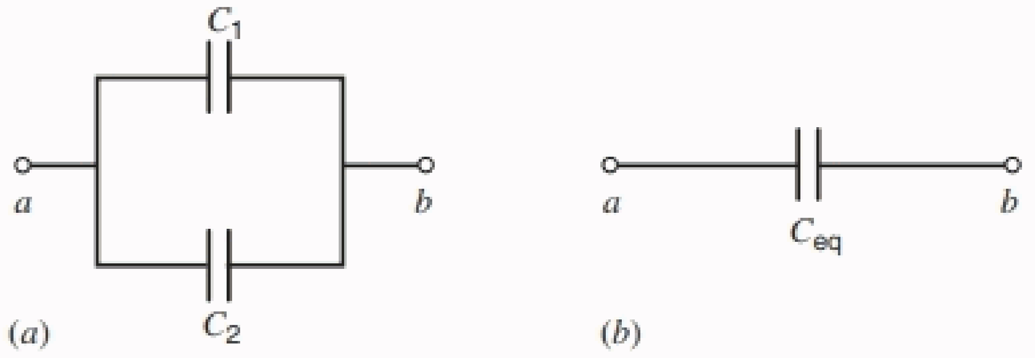


FIGURE 30-7. (*a*) Two capacitors in parallel. (*b*) The equivalent capacitance that can replace the parallel combination.

third characteristic, such that the sum of the charges on the two capacitors equals the total charge:

$$q = q_1 + q_2. \tag{30-14}$$

If the parallel combination were replaced with a single capacitor C_{eq} and connected to the same battery, the requirement that the circuit operate in identical fashion means that the same charge q must be transferred by the battery. That is, for the equivalent capacitor,

$$q = C_{\text{eq}}\Delta V. \tag{30-15}$$

Substituting Eq. 30-14 into Eq. 30-15, and then putting Eqs. 30-13 into the result, we obtain

$$C_{\text{eq}}\Delta V = C_1\Delta V + C_2\Delta V$$

or

$$C_{\text{eq}} = C_1 + C_2. \tag{30-16}$$

If we have more than two capacitors in parallel, we can first replace C_1 and C_2 with their equivalent C_{12}, determined according to Eq. 30-16. We then find the equivalent capacitance of C_{12} and the next parallel capacitor C_3. Continuing this process, we can extend Eq. 30-16 to any number of capacitors connected in parallel:

$$C_{\text{eq}} = \sum_n C_n \qquad \text{(parallel combination).} \tag{30-17}$$

That is, to find the equivalent capacitance of a parallel combination, simply add the individual capacitances. Note that the equivalent capacitance is always larger than the largest capacitance in the parallel combination. The parallel combination can store more charge than any one of the individual capacitors.

Capacitors Connected in Series

Figure 30-8 shows two capacitors connected *in series.* There are three properties that distinguish a series connection of circuit elements. (1) If we attempt to travel from a to b, we must pass through *all* the circuit elements *in succession.* (2) When a battery is connected across the combination, the potential difference ΔV of the battery equals the sum of the potential differences across each of the elements. (3) The charge q delivered to each element of the series combination has the same value.

To understand this last property, note the region of Fig. 30-8 enclosed by the dashed line. Let us assume the battery puts a charge $-q$ on the left-hand plate of C_1. Since a capacitor carries equal and opposite charges on its plates, a charge $+q$ appears on the right-hand plate of C_1. However, the H-shaped conductor enclosed by the dashed line is electrically isolated from the rest of the circuit; initially it carries no net charge, and no charge can be transferred to it. If a charge $+q$ appears on the right-hand plate of C_1, then a charge $-q$ must appear on the left-hand plate of C_2. That is, $n\ (=q/e)$ electrons move from the right-hand plate of C_1 to the left-hand plate of C_2. If there were more than two capacitors in series, a similar argument could be made across the entire line of capacitors, the result being that the left-hand plate of *every* capacitor in the series connection carries a charge q of one sign, and the right-hand plate of *every* capacitor in the series connection carries a charge of equal magnitude q and opposite sign.

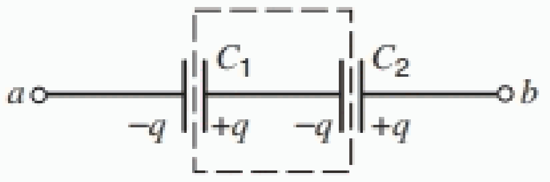


FIGURE 30-8. A series combination of two capacitors.

For the individual capacitors we can write, using Eq. 30-1,

$$\Delta V_1 = \frac{q}{C_1} \quad \text{and} \quad \Delta V_2 = \frac{q}{C_2}, \tag{30-18}$$

with the same charge q on each capacitor but different potential differences across each. According to the second property of a series connection, we have

$$\Delta V = \Delta V_1 + \Delta V_2. \tag{30-19}$$

We seek the equivalent capacitance C_{eq} that can replace the combination, such that the battery would move the same amount of charge:

$$\Delta V = \frac{q}{C_{\text{eq}}}. \tag{30-20}$$

Substituting Eq. 30-19 into Eq. 30-20 and then using Eqs. 30-18, we obtain

$$\frac{q}{C_{\text{eq}}} = \frac{q}{C_1} + \frac{q}{C_2},$$

or

$$\frac{1}{C_{\text{eq}}} = \frac{1}{C_1} + \frac{1}{C_2}. \tag{30-21}$$

If we have several capacitors in series, we can use Eq. 30-21 to find the equivalent capacitance C_{12} of the first two. We then find the equivalent capacitance of C_{12} and the next capacitor in series, C_3. Continuing in this way, we find the equivalent capacitance of any number of capacitors in series,

$$\frac{1}{C_{\text{eq}}} = \sum_n \frac{1}{C_n} \qquad \text{(series combination).} \tag{30-22}$$

That is, to find the equivalent capacitance of a series combination, take the reciprocal of the sum of the reciprocals of the individual capacitances. Note that the equivalent capacitance of the series combination is always smaller than the smallest individual capacitance in the series.

Occasionally, capacitors are connected in ways that are not immediately identifiable as series or parallel combinations. As Sample Problem 30-5 shows, such combinations can often (but not always) be broken down into smaller units that can be analyzed as series or parallel connections.

SAMPLE PROBLEM 30-5. (*a*) Find the equivalent capacitance of the combination shown in Fig. 30-9*a*, with $C_1 = 12.0\ \mu\text{F}$,

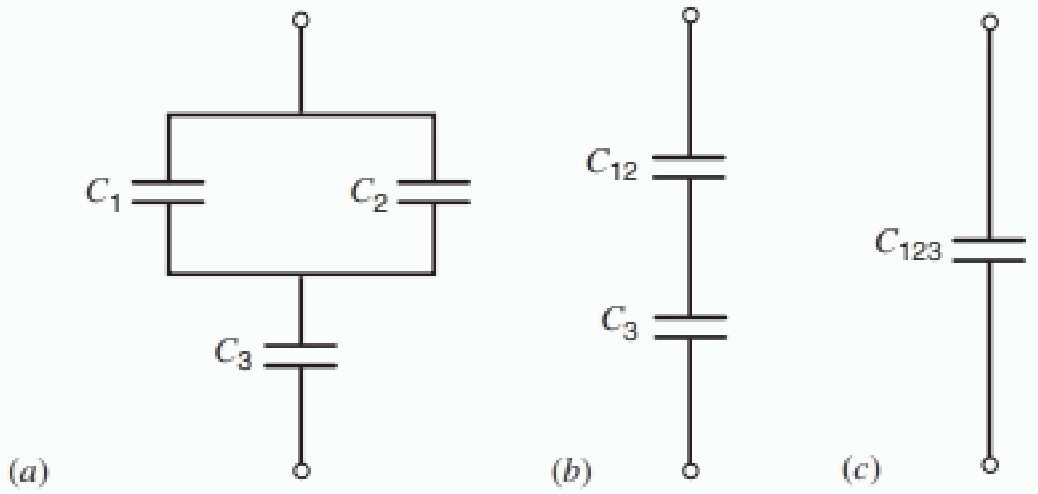


FIGURE 30-9. Sample Problem 30-5. (*a*) A combination of three capacitors. (*b*) The parallel combination of C_1 and C_2 has been replaced by its equivalent, C_{12}. (*c*) The series combination of C_{12} and C_3 has been replaced by its equivalent, C_{123}.

$C_2 = 5.3\ \mu\text{F}$, and $C_3 = 4.5\ \mu\text{F}$. (*b*) A potential difference $\Delta V = 12.5$ V is applied to the terminals in Fig. 30-9*a*. What is the charge on C_1?

Solution (*a*) Capacitors C_1 and C_2 are in parallel. From Eq. 30-16, their equivalent capacitance is

$$C_{12} = C_1 + C_2 = 12.0\ \mu\text{F} + 5.3\ \mu\text{F} = 17.3\ \mu\text{F}.$$

In Fig. 30-9*b*, C_1 and C_2 have been replaced by their parallel combination, C_{12}. As the figure shows, C_{12} and C_3 are in series. From Eq. 30-21, the final equivalent combination (see Fig. 30-9*c*) is found from

$$\frac{1}{C_{123}} = \frac{1}{C_{12}} + \frac{1}{C_3} = \frac{1}{17.3\ \mu\text{F}} + \frac{1}{4.5\ \mu\text{F}} = 0.280\ \mu\text{F}^{-1},$$

or

$$C_{123} = \frac{1}{0.280\ \mu\text{F}^{-1}} = 3.57\ \mu\text{F}.$$

(*b*) We treat the equivalent capacitors C_{12} and C_{123} exactly as we would real capacitors having that capacitance. The charge on C_{123} in Fig. 30-9*c* is then

$$q_{123} = C_{123}\,\Delta V = (3.57\ \mu\text{F})(12.5\ \text{V}) = 44.6\ \mu\text{C}.$$

This same charge exists on each capacitor in the series combination of Fig. 30-9*b*. The potential difference across C_{12} in that figure is then

$$\Delta V_{12} = \frac{q_{12}}{C_{12}} = \frac{44.6\ \mu\text{C}}{17.3\ \mu\text{F}} = 2.58\ V.$$

This same potential difference appears across C_1 in Fig. 30-9*a*, so that

$$q_1 = C_1\,\Delta V_1 = (12\ \mu\text{F})(2.58\ \text{V}) = 31\ \mu\text{C}.$$

30-5 ENERGY STORAGE IN AN ELECTRIC FIELD

An important use of capacitors is to store electrostatic energy in applications ranging from flash lamps to laser systems (see Fig. 30-10), both of which depend for their operation on the charging and discharging of capacitors.

In Section 28-2 we showed that any charge configuration has a certain *electric potential energy U*, equal to the work W (which may be positive or negative) that is done by an external agent that assembles the charge configuration from its individual components, originally assumed to be infinitely far apart and at rest. This potential energy is similar to that of mechanical systems, such as a compressed spring or the Earth–Moon system.

FIGURE 30-10. This bank of 10,000 capacitors at the Lawrence Livermore National Laboratory stores 60 MJ of electric energy and releases it in 1 ms to flashlamps that drive a system of lasers. The installation is part of the Nova project, which is attempting to produce sustained nuclear fusion reactions.

For a simple example, work is done when two equal and opposite charges are separated. This energy is stored as electric potential energy in the system, and it can be recovered as kinetic energy if the charges are allowed to come together again. Similarly, a charged capacitor has stored in it an electrical potential energy U equal to the work W done by the external agent as the capacitor is charged. This energy can be recovered if the capacitor is allowed to discharge. Alternatively, we can visualize the work of charging by imagining that an external agent pulls electrons from the positive plate and pushes them onto the negative plate, thereby bringing about the charge separation. Normally, the work of charging is done by a battery, at the expense of its store of chemical energy.

Suppose that at a time t a charge q' has already been transferred from one plate to the other. The potential difference

$\Delta V'$ between the plates at that moment is $\Delta V' = q'/C$. If an increment of charge dq' is now transferred, the resulting small change dU in the electric potential energy is, according to Eq. 28-9 ($\Delta V = \Delta U/q_0$),

$$dU = \Delta V' dq' = \frac{q'}{C}\,dq'.$$

If this process is continued until a total charge q has been transferred, the total potential energy is

$$U = \int dU = \int_0^q \frac{q'}{C}\,dq' \tag{30-23}$$

or

$$U = \frac{q^2}{2C}. \tag{30-24}$$

From the relation $q = C\,\Delta V$ we can also write this as

$$U = \tfrac{1}{2}C(\Delta V)^2 \tag{30-25}$$

Where does this energy reside? Equations 30-24 and 30-25 do not give us a direct answer, but we can determine the location of the stored energy by reasoning as follows. Suppose we have an isolated parallel-plate capacitor (that is, not connected to a battery) that carries a charge q. Without changing q, we pull the plates apart until their separation is twice as large as it was initially. According to Eq. 30-5, if the plate separation d becomes twice as large, the capacitance becomes only half as large. Equation 30-24 shows that if C becomes half as large, the stored energy doubles. Now in pulling the plates apart we have not changed the capacitor plates, so it would not be reasonable to conclude that the extra energy is stored there. What we have done is to double the volume of the space between the plates, and since the energy has also doubled it seems reasonable to conclude that this electric potential energy resides in the volume between the plates. More specifically, *the energy is stored in the electric field that is present in this region.*

In a parallel-plate capacitor, neglecting fringing, the electric field has the same value for all points between the plates. Based on our conclusion that the energy resides in the field, it follows that the *energy density u*, which is the stored energy per unit volume, should also be the same everywhere between the plates; u is given by the stored energy U divided by the volume Ad, or

$$u = \frac{U}{Ad} = \frac{\frac{1}{2}C(\Delta V)^2}{Ad}. \tag{30-26}$$

Substituting the relation $C = \epsilon_0 A/d$ (Eq. 30-5) leads to

$$u = \frac{\epsilon_0}{2}\left(\frac{\Delta V}{d}\right)^2. \tag{30-27}$$

However, $\Delta V/d$ is the electric field E, so that

$$u = \tfrac{1}{2}\epsilon_0 E^2. \tag{30-28}$$

Although we derived this equation for the special case of a parallel-plate capacitor, it is true in general. *If an electric field* $\vec{\mathbf{E}}$ *exists at any point in empty space (a vacuum), we can think of that point as the site of stored energy in amount, per unit volume, of* $\frac{1}{2}\epsilon_0 E^2$.

In general, E varies with location, so u is a function of the coordinates. For the special case of the parallel-plate capacitor, E and u do not vary with location in the region between the plates.

SAMPLE PROBLEM 30-6. A 3.55-μF capacitor C_1 is charged to a potential difference $\Delta V_0 = 6.30$ V, using a battery. The charging battery is then removed, and the capacitor is connected as in Fig. 30-11 to an uncharged 8.95-μF capacitor C_2. After the switch S is closed, charge flows from C_1 to C_2 until an equilibrium is established, with both capacitors at the same potential difference ΔV. (*a*) What is this common potential difference? (*b*) What is the energy stored in the electric field before and after the switch S in Fig. 30-11 is closed?

Solution (*a*) Electric charge must be conserved, so the original charge q_0 is shared by two capacitors, or

$$q_0 = q_1 + q_2.$$

Applying the relation $q = C\,\Delta V$ to each term yields

$$C_1\,\Delta V_0 = C_1\,\Delta V + C_2\,\Delta V,$$

or

$$\Delta V = \Delta V_0 \frac{C_1}{C_1 + C_2} = \frac{(6.30\text{ V})(3.55\ \mu\text{F})}{3.55\ \mu\text{F} + 8.95\ \mu\text{F}} = 1.79\text{ V}.$$

If we know the battery voltage ΔV_0 and the value of C_1, we can determine an unknown capacitance C_2 by measuring the value of ΔV in an arrangement similar to that of Fig. 30-11.

(*b*) The initial stored energy is

$$U_i = \tfrac{1}{2}C_1(\Delta V_0)^2 = \tfrac{1}{2}(3.55 \times 10^{-6}\text{ F})(6.30\text{ V})^2$$
$$= 7.05 \times 10^{-5}\text{ J} = 70.5\ \mu\text{J}.$$

The final energy is

$$U_f = \tfrac{1}{2}C_1(\Delta V)^2 + \tfrac{1}{2}C_2(\Delta V)^2 = \tfrac{1}{2}(C_1 + C_2)(\Delta V)^2$$
$$= \tfrac{1}{2}(3.55 \times 10^{-6}\text{ F} + 8.95 \times 10^{-6}\text{ F})(1.79\text{ V})^2$$
$$= 2.00 \times 10^{-5}\text{ J} = 20.0\ \mu\text{J}.$$

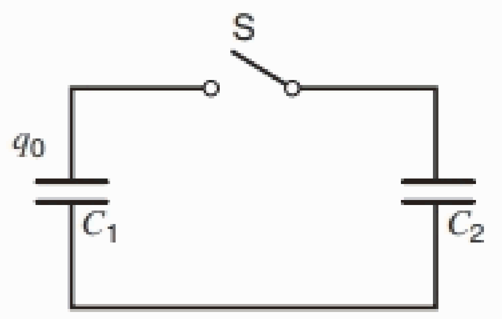


FIGURE 30-11. Sample Problem 30-6. Capacitor C_1 has previously been charged to a potential difference ΔV_0 by a battery that has been removed. When the switch S is closed, the initial charge q_0 on C_1 is shared with C_2.

We conclude that $U_f < U_i$, by about 72%. This is not a violation of energy conservation. The "missing" energy appears as thermal energy in the connecting wires, as we discuss in the next chapter.*

SAMPLE PROBLEM 30-7. An isolated conducting sphere whose radius R is 6.85 cm carries a charge $q = 1.25$ nC. (*a*) How much energy is stored in the electric field of this charged conductor? (*b*) What is the energy density at the surface of the sphere? (*c*) What is the radius R_0 of an imaginary spherical surface such that one-half of the stored potential energy lies within it?

Solution (*a*) From Eqs. 30-24 and 30-12 we have

$$U = \frac{q^2}{2C} = \frac{q^2}{8\pi\epsilon_0 R} = \frac{(1.25 \times 10^{-9}\ \mathrm{C})^2}{(8\pi)(8.85 \times 10^{-12}\ \mathrm{F/m})(0.0685\ \mathrm{m})}$$

$$= 1.03 \times 10^{-7}\ \mathrm{J} = 103\ \mathrm{nJ}.$$

(*b*) To find the energy density, we must first find E at the surface of the sphere. This is given by

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{R^2}.$$

The energy density is then, using Eq. 30-28,

$$u = \tfrac{1}{2}\epsilon_0 E^2 = \frac{q^2}{32\pi^2\epsilon_0 R^4}$$

$$= \frac{(1.25 \times 10^{-9}\ \mathrm{C})^2}{(32\pi^2)(8.85 \times 10^{-12}\ \mathrm{C^2/N\cdot m^2})(0.0685\ \mathrm{m})^4}$$

$$= 2.54 \times 10^{-5}\ \mathrm{J/m^3} = 25.4\ \mu\mathrm{J/m^3}.$$

(*c*) The energy that lies in a spherical shell between radii r and $r + dr$ is

$$dU = (u)(4\pi r^2)(dr),$$

where $(4\pi r^2)(dr)$ is the volume of the spherical shell. Using the result of part (*b*) for the energy density evaluated at a radius r, we obtain

$$dU = \frac{q^2}{32\pi^2\epsilon_0 r^4}\,4\pi r^2\,dr = \frac{q^2}{8\pi\epsilon_0}\frac{dr}{r^2}.$$

The condition given for this problem is

$$\int_R^{R_0} dU = \frac{1}{2}\int_R^{\infty} dU$$

or, using the result obtained above for dU and canceling constant factors from both sides,

$$\int_R^{R_0}\frac{dr}{r^2} = \frac{1}{2}\int_R^{\infty}\frac{dr}{r^2},$$

which becomes

$$\frac{1}{R} - \frac{1}{R_0} = \frac{1}{2R}.$$

Solving for R_0 yields

$$R_0 = 2R = (2)(6.85\ \mathrm{cm}) = 13.7\ \mathrm{cm}.$$

Half the stored energy is contained within a spherical surface whose radius is twice the radius of the conducting sphere.

30-6 CAPACITOR WITH DIELECTRIC

In Section 29-6 we discussed the effect of applying an electric field to an insulating material (a dielectric). We showed that the effect of the dielectric is to reduce the strength of the electric field in its interior from its initial value E_0 in vacuum to $E = E_0/\kappa_e$ inside the dielectric. The parameter κ_e, the dielectric constant, has values greater than 1 for all materials, so that the electric field in the dielectric is smaller than the field in vacuum.

In this section, we consider the effect of filling the interior of a capacitor with a dielectric material. This effect was first investigated in 1837 by Michael Faraday. Faraday constructed two identical capacitors, filling one with a dielectric material and leaving the other with air between its plates. When both capacitors were connected to batteries with the *same potential difference,* Faraday found that the charge on the capacitor filled with the dielectric was greater than the charge on the capacitor with air between its plates. That is, the presence of the dielectric enables the capacitor to store more charge. Since storage of charge for later discharge is one of the purposes for which we use capacitors, the presence of a dielectric can enhance the performance of a capacitor.

The effect of filling a capacitor with dielectric depends on whether we do so with the battery connected (as in Faraday's experiment) or disconnected. First we consider the situation as in Faraday's experiment (Fig. 30-12). A capacitor with capacitance C is connected to a battery of potential difference ΔV and allowed to become fully charged, such that the plates carry a charge q, as in Fig. 30-12*a*. With the battery remaining connected, we then fill the interior of the capacitor with a material of dielectric constant κ_e, as in Fig. 30-12*b*. The battery maintains the same potential difference ΔV across the plates.

Equation 30-2 shows that, if the potential differences in Figs. 30-12*a* and 30-12*b* are the same, then the electric fields inside the capacitor must be the same. However, we

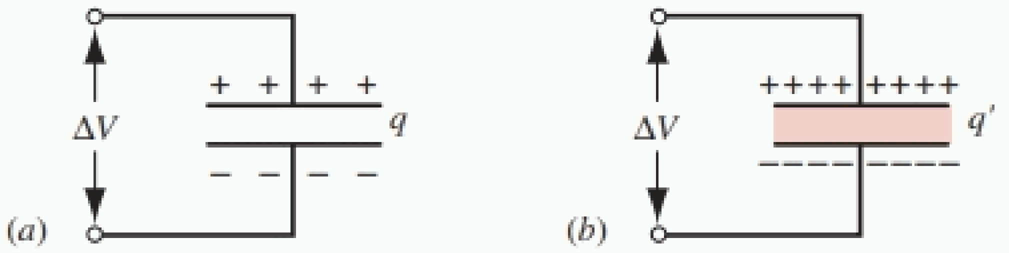


FIGURE 30-12. (*a*) An empty capacitor is charged by connecting it to a battery that establishes a potential difference ΔV. (*b*) The battery remains connected as the capacitor is filled with a dielectric. In this case, the potential difference ΔV remains constant, but q increases.

---

*Some slight amount of energy is also radiated away. For a critical discussion, see "Two-Capacitor Problem: A More Realistic View," by R. A. Powell, *American Journal of Physics,* May 1979, p. 460.

would expect the presence of the dielectric to reduce the strength of the electric field. As Faraday concluded, the tendency of the dielectric to reduce the field is exactly balanced by the additional charge that the battery delivers to the plates as the dielectric is inserted.

Let us assume that we are using a parallel-plate capacitor. With the capacitor empty, the electric field is given by Eq. 30-3: $E = \sigma/\epsilon_0 = q/\epsilon_0 A$. When the dielectric is present, the electric field is reduced by the factor $1/\kappa_e$ due to the presence of the dielectric, but the field is also changed because the plates now carry a charge q', so the field is $E' = q'/\kappa_e\epsilon_0 A$. Since the fields must be equal, we can set $E' = E$ and conclude that

$$q' = \kappa_e q. \tag{30-29}$$

The dielectric constant is greater than 1, so the capacitor can store more charge with the dielectric present than it can when it is empty. As the dielectric material is inserted into the already-charged capacitor, the battery moves additional charge $q' - q = q(\kappa_e - 1)$ from the negative to the positive plate.

The capacitance with the dielectric present is $C' = q'/\Delta V'$. Using $q' = \kappa_e q$ and $\Delta V' = \Delta V$, we obtain

$$C' = \kappa_e C. \tag{30-30}$$

The presence of the dielectric increases the capacitance by the factor κ_e. For a parallel-plate capacitor with dielectric, the capacitance can be found by combining Eqs. 30-5 and 30-30:

$$C' = \frac{\kappa_e \epsilon_0 A}{d}. \tag{30-31}$$

The capacitance of *any* capacitor is increased by the same factor κ_e when a dielectric substance completely fills the space between the plates. Equations 30-8 and 30-11 can be similarly modified to account for the presence of a dielectric filling the capacitor.

Although the effect on the capacitance is the same, the derivation is very different if the dielectric material is inserted with the battery *not* connected. We first connect the capacitor to a battery, so that the plates acquire a potential difference ΔV and charge q, after which we disconnect the battery, as in Fig. 30-13*a*. Now we fill the capacitor with dielectric, as in Fig. 30-13*b*. In this case, *the charge must remain constant,* since there is no battery present to move charge from one plate to another. With the charge remaining constant, the electric field is changed only by the presence of the dielectric, so $E' = E/\kappa_e$. Using this electric field in Eq. 30-2 to find the potential difference, we obtain $\Delta V' = \Delta V/\kappa_e$. That is, the potential difference in this case decreases by the factor $1/\kappa_e$. With $\Delta V' = q'/C'$ and $q' = q$, once again we obtain $C' = \kappa_e C$, as in Eq. 30-30. The capacitance does not depend on *how* we charge the capacitor or insert the dielectric; it depends only on the geometry of the capacitor and the material with which it is filled.

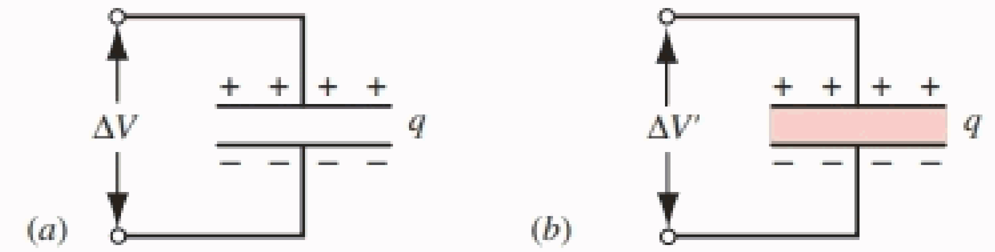


FIGURE 30-13. (*a*) An empty capacitor is charged and then disconnected from a battery. (*b*) The capacitor is now filled with a dielectric. The charge remains constant, but the potential difference decreases from ΔV to $\Delta V'$.

SAMPLE PROBLEM 30-8. A parallel-plate capacitor whose capacitance C is 13.5 pF has a potential difference $\Delta V = 12.5$ V across its plates. The charging battery is now disconnected and a porcelain slab ($\kappa_e = 6.5$) is slipped between the plates as in Fig. 30-13*b*. What is the stored energy of the unit, both before and after the slab is introduced?

Solution The initial stored energy is given by Eq. 30-25 as

$$U_i = \tfrac{1}{2}C\,\Delta V^2 = \tfrac{1}{2}(13.5 \times 10^{-12}\,\text{F})(12.5\,\text{V})^2$$
$$= 1.055 \times 10^{-9}\,\text{J} = 1055\,\text{pJ}.$$

We can write the final energy from Eq. 30-24 in the form $U_f = q^2/2C'$ because, from the conditions of the problem statement, q (but not ΔV) remains constant as the slab is introduced. After the slab is in place, the capacitance increases to $C' = \kappa_e C$ so that

$$U_f = \frac{q^2}{2\kappa_e C} = \frac{U_i}{\kappa_e} = \frac{1055\,\text{pJ}}{6.5} = 162\,\text{pJ}.$$

The energy after the slab is introduced is smaller by a factor of $1/\kappa_e$.

The "missing" energy, in principle, would be apparent to the person who introduced the slab. The capacitor would exert a force on the slab and would do work on it, in the amount

$$W = U_i - U_f = 1055\,\text{pJ} - 162\,\text{pJ} = 893\,\text{pJ}.$$

If the slab were introduced with no restraint and if there were no friction, the slab would oscillate into and out of the region between the plates. The system consisting of capacitor + slab has a constant energy of 1055 pJ; the energy shuttles back and forth between kinetic energy of the moving slab and stored energy of the electric field. At the instant the oscillating slab filled the space between the plates, its kinetic energy would be 893 pJ.

Dielectrics and Gauss' Law

So far our use of Gauss' law has been confined to situations in which no dielectric was present. Now let us apply this law to a parallel-plate capacitor filled with a material of dielectric constant κ_e.

Figure 30-14 shows the capacitor both with and without the dielectric. We assume that the charge q on the plates is the same in each case. Gaussian surfaces have been drawn partly through the upper plate and partly through the region between the plates.

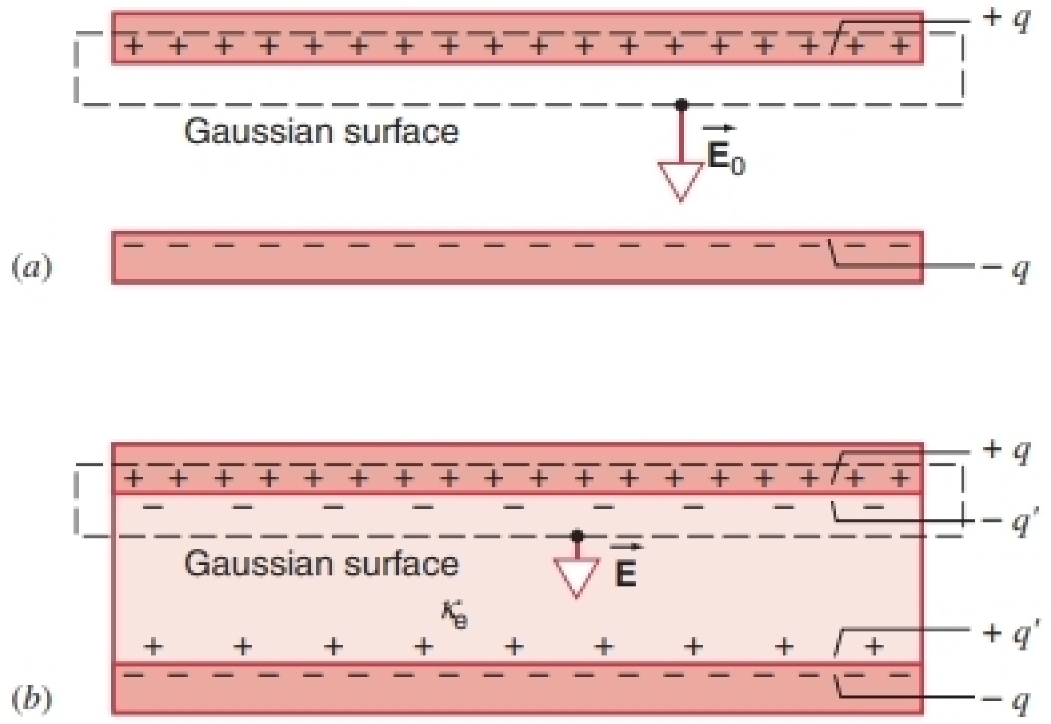


FIGURE 30-14. (*a*) A parallel-plate capacitor. (*b*) A dielectric slab is inserted, while the charge q on the plates remains constant. Induced charged q' appears on the surface of the dielectric slab.

If no dielectric is present (Fig. 30-14*a*), Gauss' law gives

$$\epsilon_0 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \epsilon_0 E_0 A = q,$$

because the electric field exists only on that portion of the Gaussian surface between the plates. Thus,

$$E_0 = \frac{q}{\epsilon_0 A}. \tag{30-32}$$

If the dielectric is present (Fig. 30-14*b*), Gauss' law gives

$$\epsilon_0 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \epsilon_0 E A = q - q'$$

or

$$E = \frac{q}{\epsilon_0 A} - \frac{q'}{\epsilon_0 A}, \tag{30-33}$$

in which $-q'$, the *induced surface charge,* must be distinguished from q, the *free charge* on the plates. These two charges $+q$ and $-q'$, both of which lie within the Gaussian surface, are opposite in sign; the *net* charge within the Gaussian surface is $q + (-q') = q - q'$.

The dielectric reduces the electric field by the factor κ_e, and so

$$E = \frac{E_0}{\kappa_e} = \frac{q}{\kappa_e \epsilon_0 A}. \tag{30-34}$$

Inserting this in Eq. 30-33 yields

$$\frac{q}{\kappa_e \epsilon_0 A} = \frac{q}{\epsilon_0 A} - \frac{q'}{\epsilon_0 A}$$

or

$$q' = q\left(1 - \frac{1}{\kappa_e}\right). \tag{30-35}$$

This shows that the induced surface charge q' is always less in magnitude than the free charge q and is equal to zero if no dielectric is present—that is, if $\kappa_e = 1$.

Now we write Gauss' law for the case of Fig. 30-14*b* in the form

$$\epsilon_0 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q - q', \tag{30-36}$$

$q - q'$ again being the net charge within the Gaussian surface. Substituting from Eq. 30-35 for q' leads, after some rearrangement, to

$$\epsilon_0 \oint \kappa_e \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q. \tag{30-37}$$

This important relation, although derived for a parallel-plate capacitor, is true generally and is the form in which Gauss' law is usually written when dielectrics are present. Note the following:

1. The flux integral now deals with $\kappa_e \vec{\mathbf{E}}$ instead of $\vec{\mathbf{E}}$. This is consistent with the *reduction* of E in a dielectric by the factor κ_e, because $\kappa_e \vec{\mathbf{E}}$ (dielectric present) equals $\vec{\mathbf{E}}_0$ (no dielectric). For generality, we allow for the possibility that the dielectric is not uniform by putting κ_e inside the integral.

2. The charge q contained within the Gaussian surface is taken to be the *free charge only.* Induced surface charge is deliberately omitted on the right side of Eq. 30-37, having been taken into account by the introduction of κ_e on the left side. Equations 30-36 and 30-37 are completely equivalent formulations.

SAMPLE PROBLEM 30-9. Figure 30-15 shows a parallel-plate capacitor of plate area A and plate separation d. A potential difference ΔV is applied across the plates. The battery is then disconnected, and a dielectric slab of thickness b and dielectric constant κ_e is placed between the plates as shown. Assume that

$$A = 115 \text{ cm}^2, \quad d = 1.24 \text{ cm}, \quad b = 0.78 \text{ cm},$$
$$\kappa_e = 2.61, \quad \Delta V = 85.5 \text{ V}.$$

(*a*) What is the capacitance C before the slab is inserted? (*b*) What free charge appears on the plates? (*c*) What is the electric

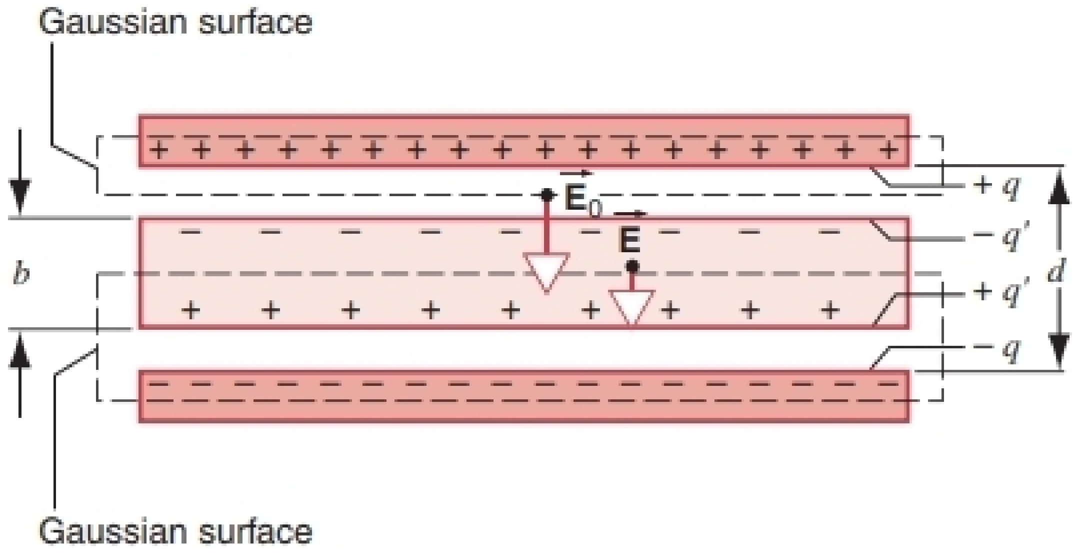


FIGURE 30-15. Sample Problem 30-9. A parallel-plate capacitor contains a dielectric that only partially fills the space between the plates.

field E_0 in the gaps between the plates and the dielectric slab? (*d*) Calculate the electric field E in the dielectric slab. (*e*) What is the potential difference $\Delta V'$ between the plates after the slab has been introduced? (*f*) What is the capacitance C' with the slab in place?

Solution (*a*) From Eq. 30-5 we have

$$C = \frac{\epsilon_0 A}{d} = \frac{(8.85 \times 10^{-12}\ \text{F/m})(115 \times 10^{-4}\ \text{m}^2)}{1.24 \times 10^{-2}\ \text{m}}$$
$$= 8.21 \times 10^{-12}\ \text{F} = 8.21\ \text{pF}.$$

(*b*) The free charge on the plates can be found from Eq. 30-1,

$$q = C\Delta V = (8.21 \times 10^{-12}\ \text{F})(85.5\ \text{V})$$
$$= 7.02 \times 10^{-10}\ \text{C} = 702\ \text{pC}.$$

Because the charging battery was disconnected before the slab was introduced, the free charge remains unchanged as the slab is put into place.

(*c*) Let us apply Gauss' law in the form given in Eq. 30-37 to the upper Gaussian surface in Fig. 30-15, which encloses only the free charge on the upper capacitor plate. We have

$$\epsilon_0 \oint \kappa_e \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \epsilon_0(1)E_0 A = q$$

or

$$E_0 = \frac{q}{\epsilon_0 A} = \frac{7.02 \times 10^{-10}\ \text{C}}{(8.85 \times 10^{-12}\ \text{F/m})(115 \times 10^{-4}\ \text{m}^2)}$$
$$= 6900\ \text{V/m} = 6.90\ \text{kV/m}.$$

Note that we put $\kappa_e = 1$ in this equation because the Gaussian surface over which Gauss' law was integrated does not pass through any dielectric. Note too that the value of E_0 remains unchanged as the slab is introduced. It depends only on the free charge on the plates.

(*d*) Again we apply Eq. 30-37, this time to the lower Gaussian surface in Fig. 30-15 and including only the free charge $-q$. We find

$$\epsilon_0 \oint \kappa_e \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = -\epsilon_0 \kappa_e E A = -q$$

or

$$E = \frac{q}{\kappa_e \epsilon_0 A} = \frac{E_0}{\kappa_e} = \frac{6.90\ \text{kV/m}}{2.61} = 2.64\ \text{kV/m}.$$

The negative sign appears when we evaluate the dot product $\vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$ because $\vec{\mathbf{E}}$ and $d\vec{\mathbf{A}}$ are in opposite directions, $d\vec{\mathbf{A}}$ always being in the direction of the *outward* normal to the closed Gaussian surface.

(*e*) To find the potential difference $\Delta V'$, we use Eq. 30-2

$$\Delta V' = \int_{+}^{-} E\, ds = E_0(d - b) + Eb$$
$$= (6900\ \text{V/m})(0.0124\ \text{m} - 0.0078\ \text{m}) + (2640\ \text{V/m})(0.0078\ \text{m})$$
$$= 52.3\ \text{V}.$$

This contrasts with the original applied potential difference of 85.5 V.

(*f*) From Eq. 30-1, the capacitance with the slab in place is

$$C' = \frac{q}{\Delta V'} = \frac{7.02 \times 10^{-10}\ \text{C}}{52.3\ \text{V}}$$
$$= 1.34 \times 10^{-11}\ \text{F} = 13.4\ \text{pF}.$$

Table 30-1 summarizes the results of this sample problem and also includes the results that would have followed if the dielectric slab had completely filled the space between the plates.

TABLE 30-1 **Summary of Results for Sample Problem 30-9**

Quantity	*Unit*	*No Slab*	*Partial Slab*	*Full Slab*
C	pF	8.21	13.4	21.4
q	pC	702	702	702
q'	pC	—	433	433
ΔV	V	85.5	52.3	32.8
E_0	kV/m	6.90	6.90	6.90[a]
E	kV/m	—	2.64	2.64

[a] Assumes that a very narrow gap is present.

MULTIPLE CHOICE

30-1 Capacitors

1. Two parallel metal plates have charges q_1 and q_2. Is this an example of a capacitor?
   (A) Yes
   (B) Only if $q_1 = -q_2$
   (C) Only if the signs of q_1 and q_2 are different
   (D) No

30-2 Capacitance

2. The centers of two identical conducting spheres of radius r are separated by a distance $d > 2r$. A charge $+q$ is on one sphere, and a charge $-q$ is on the other. The capacitance of the system is C_0. Additional charge is now transferred so that the charge on each sphere is doubled.

   (*a*) What is the new capacitance C' now that the charges have changed?
   (A) $C' = 4C_0$ (B) $C' = 2C_0$
   (C) $C' = C_0$ (D) $C' = C_0/2$
   (E) There is not enough information to answer the question.

   (*b*) What is the new potential difference $\Delta V'$ between the spheres?
   (A) $\Delta V' = 4q/C_0$ (B) $\Delta V' = 2q/C_0$
   (C) $\Delta V' = q/C_0$ (D) $\Delta V' = q/2C_0$
   (E) There is not enough information to answer the question.

3. The centers of two identical conducting spheres of radius r are separated by a distance $d > 2r$.

   (*a*) How does the capacitance of this system change if the separation of the two spheres is decreased?

(A) C increases. (B) C decreases.
(C) C remains the same.
(D) There is not enough information to answer the question.

(*b*) How does the capacitance of this system change if r is decreased?

(A) C increases. (B) C decreases.
(C) C remains the same.
(D) There is not enough information to answer the question.

30-3 Calculating the Capacitance

4. Which of the following changes to an ideal parallel-plate capacitor connected to an ideal battery will result in an increase in the charge on the capacitor?

(A) Decreasing the potential difference across the plates
(B) Decreasing the area of the plates
(C) Decreasing the separation of the plates
(D) None of the above

5. Equation 30-5 neglects fringing effects near the edge of the plates. Does this cause Eq. 30-5 to under- or overestimate the capacitance of a real parallel-plate capacitor?

(A) Overestimate (B) Underestimate
(C) Neither, the expression is exact

6. What is the capacitance of a *single* spherical conductor of radius r?

(A) $4\pi\epsilon_0$ (B) $4\pi\epsilon_0 r$ (C) $4\pi\epsilon_0/r$
(D) Capacitance is undefined for a single object.

30-4 Capacitors in Series and Parallel

7. Two capacitors C_1 and C_2 are connected in series; assume that $C_1 < C_2$. The equivalent capacitance of this arrangement is C, where

(A) $C < C_1/2$. (B) $C_1/2 < C < C_1$.
(C) $C_1 < C < C_2$. (D) $C_2 < C < 2C_2$.
(E) $2C_2 < C$.

8. Two capacitors C_1 and C_2 are connected in parallel; assume that $C_1 < C_2$. The equivalent capacitance of this arrangement is C, where

(A) $C < C_1/2$. (B) $C_1/2 < C < C_1$.
(C) $C_1 < C < C_2$. (D) $C_2 < C < 2C_2$.
(E) $2C_2 < C$.

9. Four possible capacitor arrangements are shown in Fig. 30-16 for three identical capacitors.

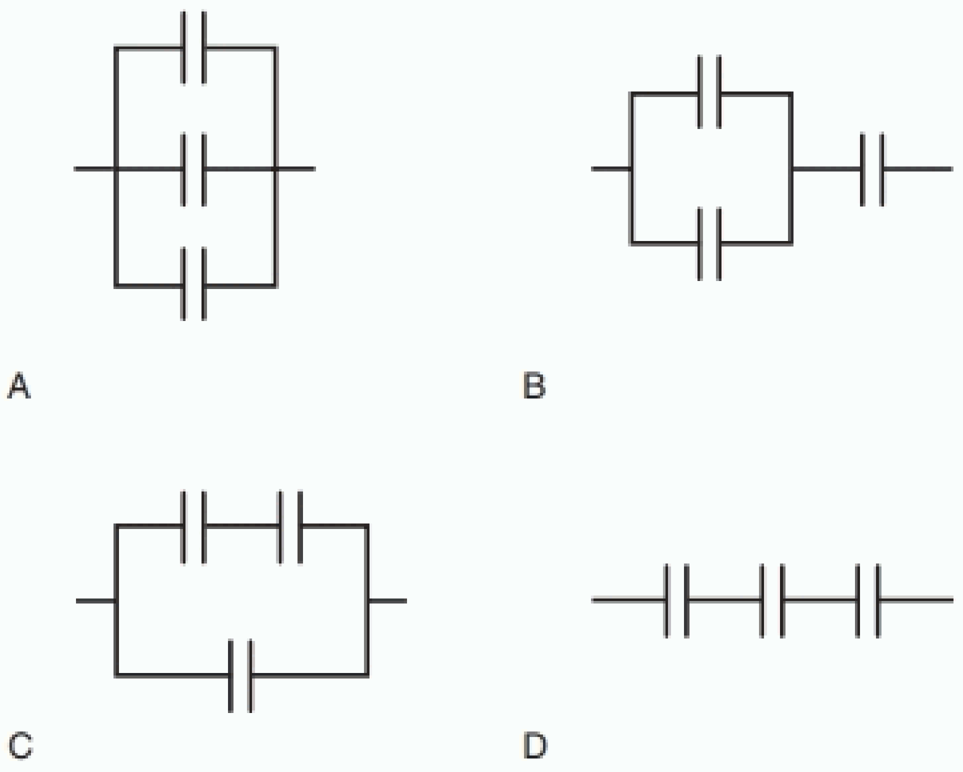


FIGURE 30-16. Multiple-choice question 9.

(*a*) Which arrangement would have the largest equivalent capacitance?

(*b*) If each arrangement were connected to a 12-V potential difference, in which case would the largest amount of charge be transferred?

(*c*) Each arrangement is connected to a potential difference so that the same amount of charge is transferred. Which requires the largest potential difference?

30-5 Energy Storage in an Electric Field

10. A parallel-plate capacitor is connected to an ideal battery, which provides a fixed potential difference. Originally the energy stored in the capacitor is U_0. If the distance between the plates is doubled, then the new energy stored in the capacitor will be

(A) $4U_0$. (B) $2U_0$. (C) U_0.
(D) $U_0/2$. (E) $U_0/4$.

11. A parallel-plate capacitor is charged by connecting it to an ideal battery; the capacitor is then disconnected. Originally the energy stored in the capacitor is U_0. If the distance between the plates is doubled, then the new energy stored in the capacitor will be

(A) $4U_0$. (B) $2U_0$. (C) U_0.
(D) $U_0/2$. (E) $U_0/4$.

12. A student originally charges a fixed capacitor to have a potential energy of 1 J. If the student wishes to give the capacitor a potential energy of 4 J, then the student should

(A) quadruple the potential difference across the capacitor but leave the charge unchanged.
(B) double the potential difference across the capacitor but leave the charge unchanged.
(C) double both the potential difference across the capacitor and the charge.
(D) leave the potential difference across the capacitor unchanged while doubling the charge.

13. An inflated balloon is covered with a conducting surface that carries a charge q. The balloon develops a leak and the radius starts to decrease, but no charge is lost from the surface. (*a*) How does the capacitance of the balloon change as the balloon leaks?

(A) C increases. (B) C decreases.
(C) C remains the same.
(D) There is not enough information to answer the question.

(*b*) How does the stored electrical energy change as the balloon leaks?

(A) U increases. (B) U decreases.
(C) U remains the same.
(D) There is not enough information to answer the question.

30-6 Capacitor with Dielectric

14. Consider a parallel-plate capacitor originally with a charge q_0, capacitance C_0, and potential difference ΔV_0. There is an electrostatic force of magnitude F_0 between the plates, and the capacitor has a stored energy U_0. The terminals of the capacitor are *not* connected to anything.

(*a*) A dielectric slab with $\kappa_e > 1$ is inserted between the plates. Which quantities increase? (Choose all that apply.)

(A) q (B) C (C) ΔV (D) F (E) U

(*b*) What is the direction of the electrostatic force on the dielectric slab while it is being inserted?

(A) The force pulls the slab into the capacitor.
(B) The force pushes the slab out of the capacitor.
(C) There is no electrostatic force on the slab.

(*c*) Later the dielectric slab is removed. What is the direction of the electrostatic force on the dielectric slab while it is being removed?

(A) The force pulls the slab into the capacitor.
(B) The force pushes the slab out of the capacitor.
(C) There is no electrostatic force on the slab.

15. Consider a parallel-plate capacitor originally with charge q_0 and capacitance C_0. There is an electrostatic force of magnitude F_0 between the plates, and the capacitor has a stored energy U_0. The terminals of the capacitor are connected to an ideal battery, which supplies a potential difference ΔV_0.

(*a*) A dielectric slab with $\kappa_e > 1$ is inserted between the plates. Which quantities increase? (Choose all that apply.)

(A) q (B) C (C) ΔV
(D) F (E) U

(*b*) What is the direction of the electrostatic force on the dielectric slab while it is being inserted?

(A) The force pulls the slab into the capacitor.
(B) The force pushes the slab out of the capacitor.
(C) There is no electrostatic force on the slab.

(*c*) Later the dielectric slab is removed. What is the direction of the electrostatic force on the dielectric slab while it is being removed?

(A) The force pulls the slab into the capacitor.
(B) The force pushes the slab out of the capacitor.
(C) There is no electrostatic force on the slab.

QUESTIONS

1. A capacitor is connected across a battery. (*a*) Why does each plate receive a charge of exactly the same magnitude? (*b*) Is this true even if the plates are of different sizes?

2. You are given two capacitors, C_1 and C_2, in which $C_1 > C_2$. How could things be arranged so that C_2 could hold more charge than C_1?

3. The relation $\sigma \propto 1/R$, in which σ is the surface charge density and R is the radius of curvature (see Eq. 28-42), suggests that the charge placed on an isolated conductor concentrates on points and avoids flat surfaces, where $R = \infty$. How do we reconcile this with Fig. 30-5, in which the charge is definitely on the flat surface of either plate?

4. In connection with Eq. 30-1 ($q = C\,\Delta V$) we said that C is a constant. Yet we pointed out (see Eq. 30-5) that it depends on the geometry (and also, as we saw later, on the medium). If C is indeed a constant, with respect to what variables does it remain constant?

5. In Fig. 30-1, suppose that a and b are nonconductors, the charge being distributed arbitrarily over their surfaces. (*a*) Would Eq. 30-1 ($q = C\,\Delta V$) hold, with C independent of the charge arrangements? (*b*) How would you define ΔV in this case?

6. You are given a parallel-plate capacitor with square plates of area A and separation d, in a vacuum. What is the qualitative effect of each of the following on its capacitance? (*a*) Reduce d. (*b*) Put a slab of copper between the plates, touching neither plate. (*c*) Double the area of both plates. (*d*) Double the area of one plate only. (*e*) Slide the plates parallel to each other so that the area of overlap is 50%. (*f*) Double the potential difference between the plates. (*g*) Tilt one plate so that the separation remains d at one end but is $\frac{1}{2}d$ at the other.

7. You have two isolated conductors, each of which has a certain capacitance; see Fig. 30-17. If you join these conductors by a fine wire, how do you calculate the capacitance of the combination? In joining them with the wire, have you connected them in series or parallel?

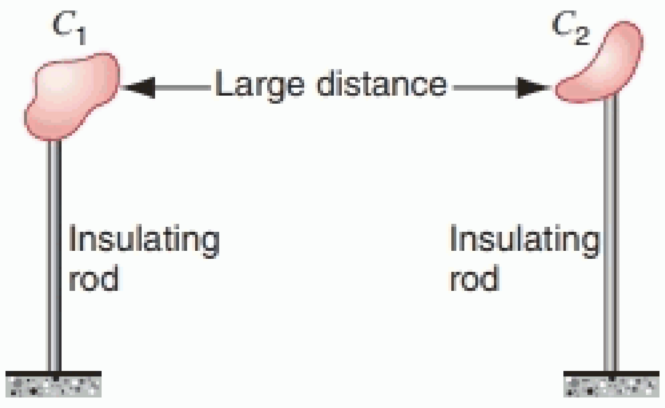


FIGURE 30-17. Question 7.

8. The capacitance of a conductor is affected by the presence of a second conductor that is uncharged and isolated electrically. Why?

9. A sheet of aluminum foil of negligible thickness is placed between the plates of a capacitor as in Fig. 30-18. What effect has it on the capacitance if (*a*) the foil is electrically insulated and (*b*) the foil is connected to the upper plate?

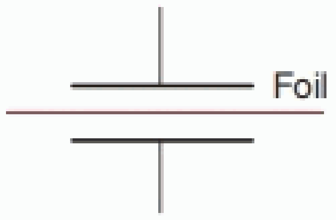


FIGURE 30-18. Question 9.

10. Capacitors often are stored with a wire connected across their terminals. Why is this done?

11. If you were not to neglect the fringing of the electric field lines in a parallel-plate capacitor, would you calculate a higher or a lower capacitance?

12. Two circular copper disks are facing each other a certain distance apart. In what ways could you reduce the capacitance of this combination?

13. Discuss similarities and differences when (*a*) a dielectric slab and (*b*) a conducting slab are inserted between the plates of a parallel-plate capacitor. Assume the slab thicknesses to be one-half the plate separation.

14. An oil-filled, parallel-plate capacitor has been designed to have a capacitance C and to operate safely at or below a certain maximum potential difference ΔV_m without arcing over. However, the designer did not do a good job and the capacitor occasionally arcs over. What can be done to redesign the capacitor, keeping C and ΔV_m unchanged and using the same dielectric?

15. For a given potential difference, does a capacitor store more or less charge with a dielectric than it does without a dielectric (vacuum)? Explain in terms of the microscopic picture of the situation.

16. Water has a high dielectric constant (see Table 29-2). Why is it not used ordinarily as a dielectric material in capacitors?

17. Figure 30-19 shows an actual 1-F capacitor available for use in student laboratories. It is only a few centimeters in diameter. Considering the result of Sample Problem 30-2, how can such a capacitor be constructed?

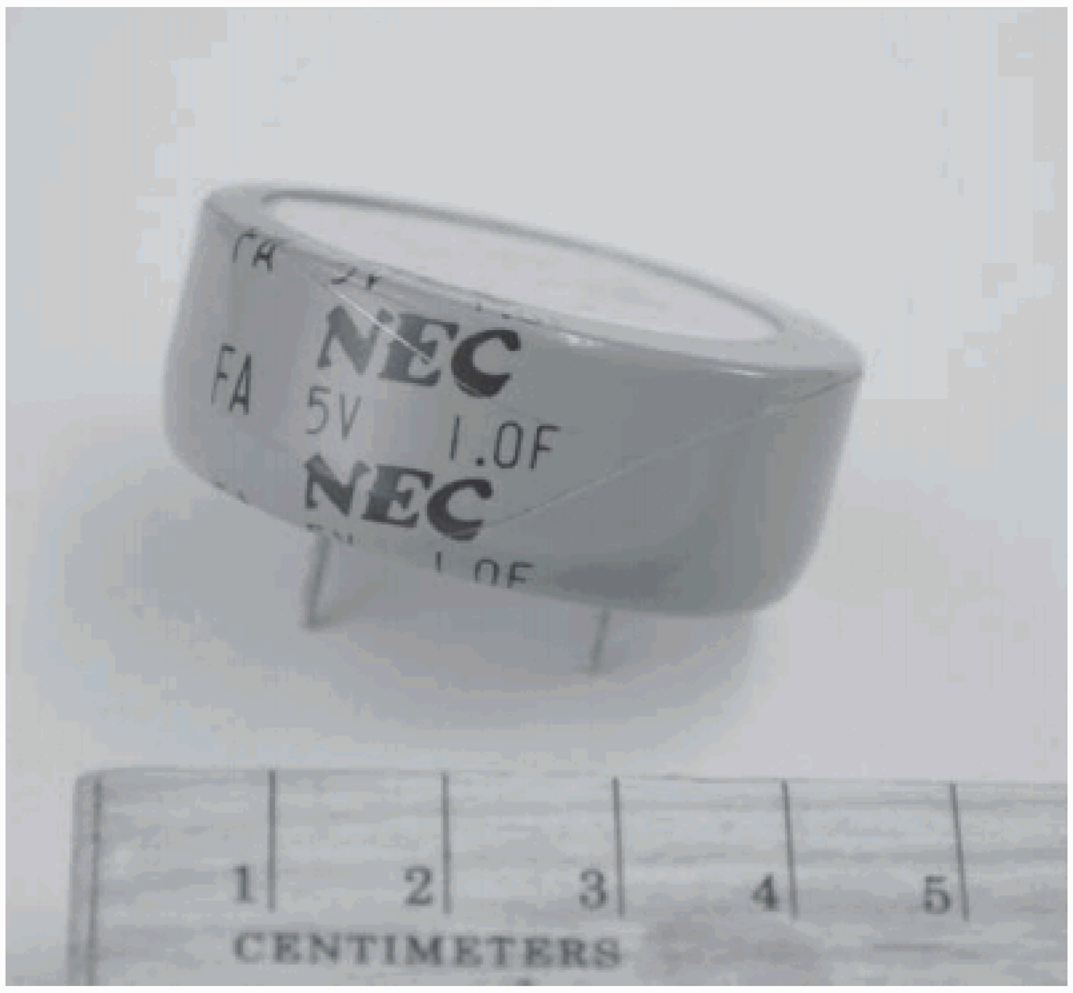


FIGURE 30-19. Question 17.

18. A dielectric slab is inserted in one end of a charged parallel-plate capacitor (the plates being horizontal and the charging battery having been disconnected) and then released. Describe what happens. Neglect friction.

19. A parallel-plate capacitor is charged by using a battery, which is then disconnected. A dielectric slab is then slipped between the plates. Describe qualitatively what happens to the charge, the capacitance, the potential difference, the electric field, and the stored energy.

20. While a parallel-plate capacitor remains connected to a battery, a dielectric slab is slipped between the plates. Describe qualitatively what happens to the charge, the capacitance, the potential difference, the electric field, and the stored energy. Is work required to insert the slab?

21. Imagine a dielectric slab, of width equal to the plate separation, inserted only halfway into a parallel-plate capacitor carrying a fixed charge q. Sketch qualitatively the distribution of the charge q on the plates and the induced charge q' on the slab.

22. Two identical capacitors are connected as shown in Fig. 30-20. A dielectric slab is slipped between the plates of one capacitor, the battery remaining connected so that a constant potential difference ΔV is maintained. Describe qualitatively what happens to the charge, the capacitance, the potential difference, the electric field, and the stored energy for each capacitor.

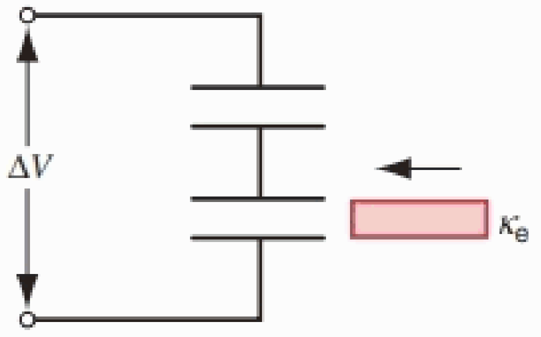


FIGURE 30-20. Question 22.

23. In this chapter we have assumed electrostatic conditions; that is, the potential difference ΔV between the capacitor plates remains constant. Suppose, however, that, as it often does in practice, ΔV varies sinusoidally with time with an angular frequency ω. Would you expect the dielectric constant κ_e to vary with ω?

EXERCISES

30-1 Capacitors

30-2 Capacitance

1. An electrometer is a device used to measure static charge. Unknown charge is placed on the plates of a capacitor and the potential difference is measured. What minimum charge can be measured by an electrometer with a capacitance of 50 pF and a voltage sensitivity of 0.15 V?

2. The two metal objects in Fig. 30-21 have net charges of +73.0 pC and −73.0 pC, and this results in a 19.2-V potential difference between them. (*a*) What is the capacitance of the system? (*b*) If the charges are changed to +210 pC and −210 pC, what does the capacitance become? (*c*) What does the potential difference become?

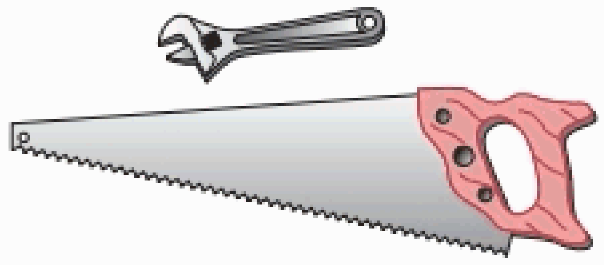


FIGURE 30-21. Exercise 2.

3. The capacitor in Fig. 30-22 has a capacitance of 26.0 μF and is initially uncharged. A battery supplies a potential difference ΔV of 125 V. After switch S has been closed for a long time, how much charge will have been moved by the battery?

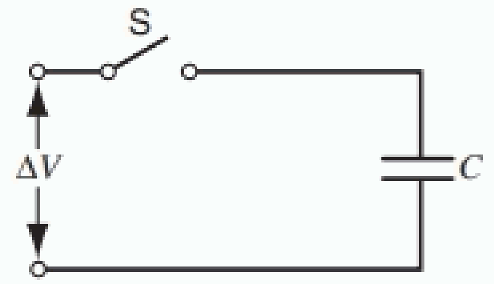


FIGURE 30-22. Exercise 3.

30-3 Calculating the Capacitance

4. A parallel-plate capacitor has circular plates of 8.22-cm radius and 1.31-mm separation. (*a*) Calculate the capacitance. (*b*) What charge will appear on the plates if a potential difference of 116 V is applied?

5. The plate and cathode of a vacuum tube diode are in the form of two concentric cylinders with the cathode as the central cylinder. The cathode diameter is 1.62 mm and the plate diameter is 18.3 mm, with both elements having a length of 2.38 cm. Calculate the capacitance of the diode.

6. Two sheets of aluminum foil have a separation of 1.20 mm, a capacitance of 9.70 pF, and are charged to 13.0 V. (*a*) Calculate the plate area. (*b*) The separation is now decreased by 0.10 mm with the charge held constant. Find the new capacitance. (*c*) By how much does the potential difference change? Explain how a microphone might be constructed using this principle.

7. The plates of a spherical capacitor have radii 38.0 mm and 40.0 mm. (*a*) Calculate the capacitance. (*b*) What must be the plate area of a parallel-plate capacitor with the same plate separation and capacitance?

8. Suppose that the two spherical shells of a spherical capacitor have their radii approximately equal. Under these conditions the device approximates a parallel-plate capacitor with $b - a = d$. Show that Eq. 30-8 for the spherical capacitor does indeed reduce to Eq. 30-5 for the parallel-plate capacitor in this case.

30-4 Capacitors in Series and Parallel

9. How many 1.00-μF capacitors must be connected in parallel to store a charge of 1.00 C with a potential difference of 110 V across the capacitors?

10. In Fig. 30-23, find the equivalent capacitance of the combination. Assume that $C_1 = 10.3\ \mu$F, $C_2 = 4.80\ \mu$F, and $C_3 = 3.90\ \mu$F.

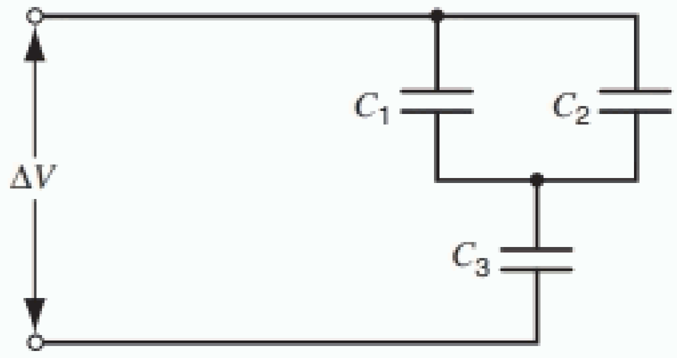


FIGURE 30-23. Exercises 10, 17, and 26.

11. In Fig. 30-24, find the equivalent capacitance of the combination. Assume that $C_1 = 10.3\ \mu$F, $C_2 = 4.80\ \mu$F, and $C_3 = 3.90\ \mu$F.

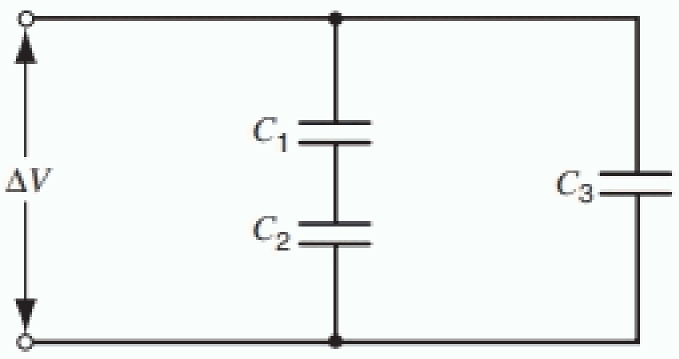


FIGURE 30-24. Exercise 11.

12. Each of the uncharged capacitors in Fig. 30-25 has a capacitance of 25.0 μF. A potential difference of 4200 V is established when the switch S is closed. How much charge then passes through the meter A?

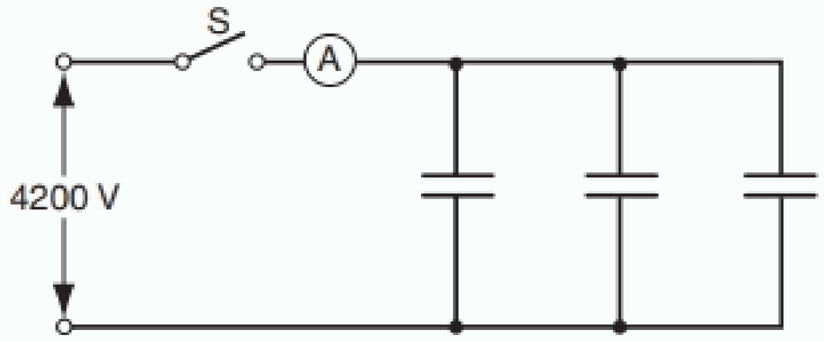


FIGURE 30-25. Exercise 12.

13. A 6.0-μF capacitor is connected in series with a 4.0-μF capacitor; a potential difference of 200 V is applied across the pair. (*a*) Calculate the equivalent capacitance. (*b*) What is the charge on each capacitor? (*c*) What is the potential difference across each capacitor?

14. Work Exercise 13 for the same two capacitors connected in parallel.

15. (*a*) Three capacitors are connected in parallel. Each has plate area *A* and plate spacing *d*. What must be the spacing of a single capacitor of plate area *A* if its capacitance equals that of the parallel combination? (*b*) What must be the spacing if the three capacitors are connected in series?

16. You have several 2.0-μF capacitors, each capable of withstanding 200 V without breakdown. How would you assemble a combination having an equivalent capacitance of (*a*) 0.40 μF or (*b*) 1.2 μF, each combination capable of withstanding 1000 V?

17. In Fig. 30-23, suppose that capacitor C_3 breaks down electrically, becoming equivalent to a conducting path. What changes in (*a*) the charge and (*b*) the potential difference occur for capacitor C_1? Assume that $\Delta V = 115$ V.

18. A 108-μF capacitor is charged to a potential difference of 52.4 V, then the charging battery is disconnected. The capacitor is then connected in parallel with a second (initially uncharged) capacitor. The measured potential difference drops to 35.8 V. Find the capacitance of this second capacitor.

19. A portion of an infinite array of identical 1 μF-capacitors is shown in Fig. 30-26. The array is initially uncharged. A battery is then connected across two distant junctions. Show that the potential at any junction is the average of the potential at the four nearest junctions. You will use the result to solve Computer Problem 1. (Hint: What is the net charge on any junction?)

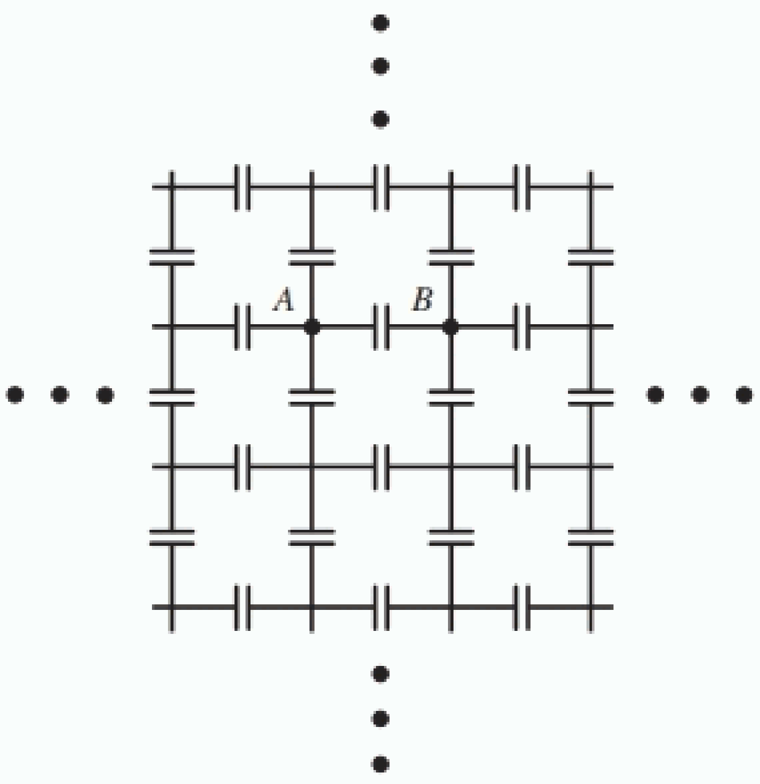


FIGURE 30-26. Exercise 19 and computer problem 1.

30-5 Energy Storage in an Electric Field

20. How much energy is stored in 2.0 m^3 of air due to the "fair weather" electric field of strength 150 V/m?

21. A parallel-connected bank of 2100 5.0-μF capacitors is used to store electric energy. What does it cost to charge this bank to 55 kV, assuming a rate of 3.0¢/kW · h?

22. Attempts to build a controlled thermonuclear fusion reactor, which, if successful, could provide the world with a vast supply of energy from heavy hydrogen in seawater, usually involve huge electric currents for short periods of time in magnetic field windings. These currents are often provided by discharging large banks of capacitors. One such capacitor bank provides 61.0 mF at 10.0 kV. Calculate the stored energy (*a*) in joules and (*b*) in kW · h.

23. A parallel-plate, air-filled capacitor having area 42.0 cm^2 and spacing of 1.30 mm is charged to a potential difference of 625 V. Find (*a*) the capacitance, (*b*) the magnitude of the charge on each plate, (*c*) the stored energy, (*d*) the electric field between the plates, and (*e*) the energy density between the plates.

24. Two capacitors, 2.12 μF and 3.88 μF, are connected in series across a 328-V potential difference. Calculate the total energy stored in the capacitors.

25. An isolated metal sphere whose diameter is 12.6 cm has a potential of 8150 V (where $V = 0$ at infinity). Calculate the energy density in the electric field near the surface of the sphere.

26. In Fig. 30-23, find (*a*) the charge, (*b*) the potential difference, and (*c*) the stored energy for each capacitor. Assume the numerical values of Exercise 10, with $\Delta V = 112$ V.

27. A cylindrical capacitor has radii a and b as in Fig. 30-6. Show that half the stored electric potential energy lies within a cylinder whose radius is $r = \sqrt{ab}$.

30-6 Capacitor with Dielectric

28. An air-filled, parallel-plate capacitor has a capacitance of 51.3 pF. (*a*) If its plates each have an area of 0.350 m^2, what is their separation? (*b*) If the region between the plates is now filled with material having a dielectric constant of 5.60, what is the capacitance?

29. An air-filled, parallel-plate capacitor has a capacitance of 1.32 pF. The separation of the plates is doubled and wax is inserted between them. The new capacitance is 2.57 pF. Find the dielectric constant of the wax.

30. Given a 7.40-pF air-filled capacitor, you are asked to design a capacitor to store up to 6.61 μJ with a maximum potential difference of 630 V. What dielectric in Table 29-2 will you use to fill the gap in the capacitor if you do not allow for a margin of error?

31. For making a parallel-plate capacitor you have available two plates of copper, a sheet of mica (thickness = 0.10 mm, $\kappa_e = 5.4$), a sheet of glass (thickness = 0.20 mm, $\kappa_e = 7.0$), and a slab of paraffin (thickness = 1.0 cm, $\kappa_e = 2.0$). To obtain the largest capacitance, which sheet should you place between the copper plates?

32. A certain substance has a dielectric constant of 2.80 and a dielectric strength of 18.2 MV/m. If the substance is used as the dielectric material in a parallel-plate capacitor, what minimum area may the plates of the capacitor have in order that the capacitance be 68.4 nF and that the capacitor be able to withstand a potential difference of 4.13 kV?

33. A coaxial cable used in a transmission line responds as a "distributed" capacitance to the circuit feeding it. Calculate the capacitance of 1.00 km for a cable having an inner radius of 0.110 mm and an outer radius of 0.588 mm. Assume that the space between the conductors is filled with polystyrene.

34. You have been assigned to design a transportable capacitor that can store 250 kJ of energy. You select a parallel-plate type with dielectric. (*a*) What is the minimum capacitor volume achievable using a dielectric selected from those listed in Table 29-2 that have values of dielectric strength? (*b*) Modern, high-performance capacitors that can store 250 kJ have volumes of 0.087 m^3. Assuming that the dielectric used has the same dielectric strength as in (*a*), what must be its dielectric constant?

35. You are asked to construct a capacitor having a capacitance near 1000 pF and a breakdown potential in excess of 10 kV. You think of using the sides of a tall drinking glass (Pyrex), lining the inside and outside with aluminum foil (neglect the ends). What are (*a*) the capacitance and (*b*) the breakdown potential? You use a glass 15 cm tall with an inner radius of 3.6 cm and an outer radius of 3.8 cm.

36. In Sample Problem 30-8, suppose that the battery remains connected during the time that the dielectric slab is being introduced. Calculate (*a*) the capacitance, (*b*) the charge on the capacitor plates, (*c*) the electric field in the gap, and (*d*) the electric field in the slab, after the slab is introduced.

37. A slab of copper of thickness b is thrust into a parallel-plate capacitor as shown in Fig. 30-27. (*a*) What is the capacitance after the slab is introduced? (*b*) If a charge q is maintained on the plates, find the ratio of the stored energy before to that after the slab is inserted. (*c*) How much work is done on the slab as it is inserted? Is the slab pulled in or do you have to push it in?

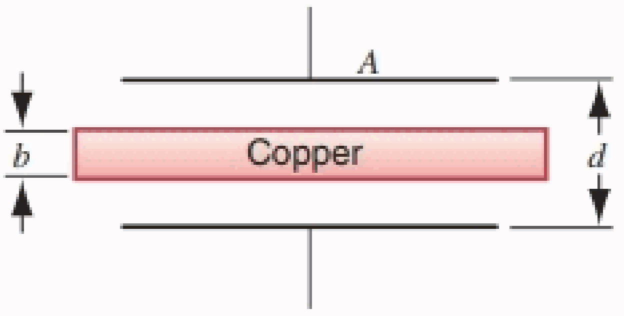


FIGURE 30-27. Exercise 37.

PROBLEMS

1. In Section 30-2, the capacitance of a cylindrical capacitor was calculated. Using the approximation (see Appendix I) that $\ln(1 + x) \approx x$ when $x \ll 1$, show that the capacitance approaches that of a parallel-plate capacitor when the spacing between the two cylinders is small.

2. A capacitor is to be designed to operate, with constant capacitance, in an environment of fluctuating temperature. As shown in Fig. 30-28, the capacitor is a parallel-plate type with plastic "spacers" to keep the plates aligned. (*a*) Show that the rate of change of capacitance C with temperature T is given by

$$\frac{dC}{dT} = C\left(\frac{1}{A}\frac{dA}{dT} - \frac{1}{x}\frac{dx}{dT}\right),$$

where A is the plate area and x is the plate separation. (*b*) If the plates are aluminum, what should be the coefficient of thermal expansion of the spacers in order that the capacitance not vary with temperature? (Ignore the effect of the spacers on the capacitance.)

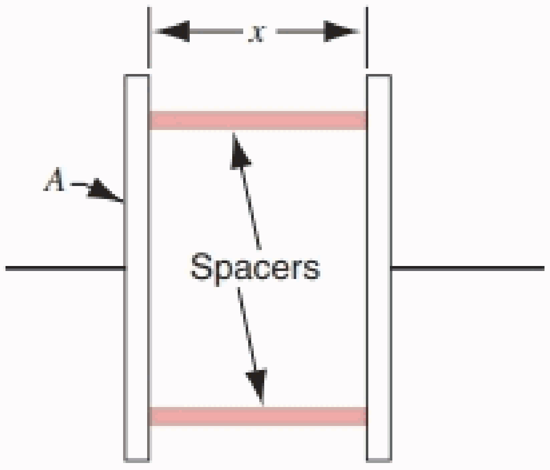


FIGURE 30-28. Problem 2.

3. Figure 30-29 shows two capacitors in series, the rigid center section of length b being movable vertically. Show that the equivalent capacitance of the series combination is independent of the position of the center section and is given by

$$C = \frac{\epsilon_0 A}{a - b}.$$

38. Reconsider Exercise 37, assuming that the potential difference ΔV rather than the charge is held constant.

39. A parallel-plate capacitor has a capacitance of 112 pF, a plate area of 96.5 cm^2, and a mica dielectric ($\kappa_e = 5.40$). At a 55.0-V potential difference, calculate the magnitudes of (*a*) the electric field in the mica, (*b*) the free charge on the plates, and (*c*) the induced surface charge.

40. Two parallel plates of area 110 cm^2 are each given equal but opposite charges of 890 nC. The electric field within the dielectric material filling the space between the plates is 1.40 MV/m. (*a*) Calculate the dielectric constant of the material. (*b*) Determine the magnitude of the charge induced on each surface of the dielectric.

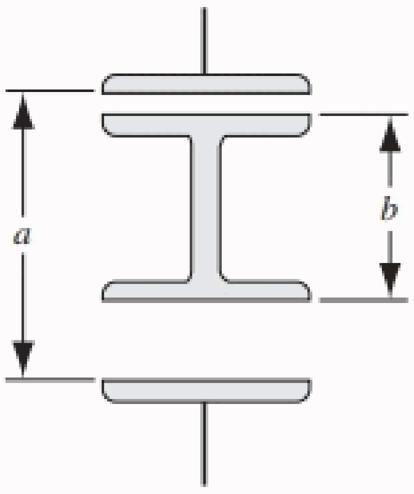


FIGURE 30-29. Problem 3.

4. In Fig. 30-30, a variable, air-filled capacitor of the type used in tuning radios is shown. Alternate plates are connected together, one group being fixed in position, the other group being capable of rotation. Consider a pile of n plates of alternate polarity, each having an area A and separated from adjacent plates by a distance d. Show that this capacitor has a maximum capacitance of

$$C = \frac{(n - 1)\epsilon_0 A}{d}.$$

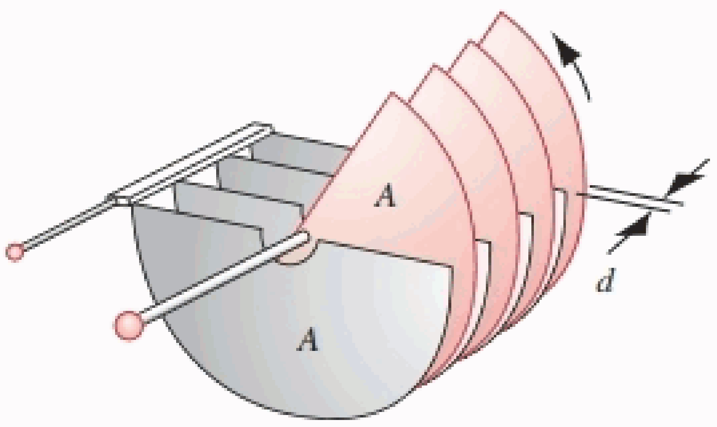


FIGURE 30-30. Problem 4.

5. In Fig. 30-31, capacitors $C_1 = 1.16\ \mu\text{F}$ and $C_2 = 3.22\ \mu\text{F}$ are each charged to a potential difference $\Delta V = 96.6$ V but with opposite polarity, so that points a and c are on the side of the respective positive plates of C_1 and C_2, and points b and d are on the side of the respective negative plates. Switches S_1 and

S_2 are now closed. (*a*) What is the potential difference between points *e* and *f*? (*b*) What is the charge on C_1? (*c*) What is the charge on C_2?

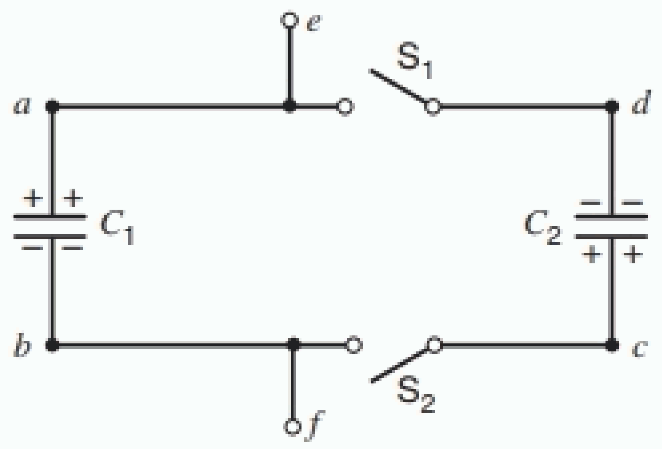


FIGURE 30-31. Problem 5.

6. When switch S is thrown to the left in Fig. 30-32, the plates of the capacitor C_1 acquire a potential difference ΔV_0. C_2 and C_3 are initially uncharged. The switch is now thrown to the right. What are the final charges q_1, q_2, q_3 on the corresponding capacitors?

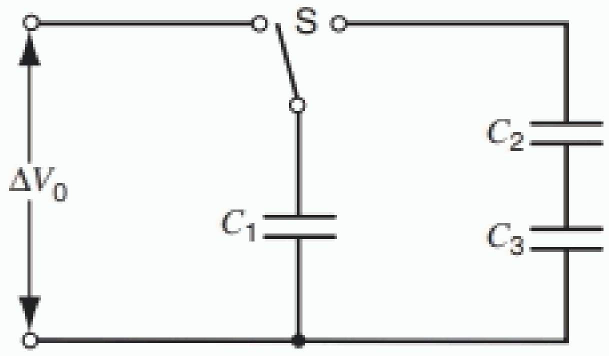


FIGURE 30-32. Problem 6.

7. Figure 30-33 shows two identical capacitors of capacitance C in a circuit with two (ideal) diodes D. A 100-V battery is connected to the input terminals, (*a*) first with terminal *a* positive and (*b*) later with terminal *b* positive. In each case, what is the potential difference across the output terminals? (An ideal diode has the property that positive charge flows through it only in the direction of the arrow and negative charge flows through it only in the opposite direction.)

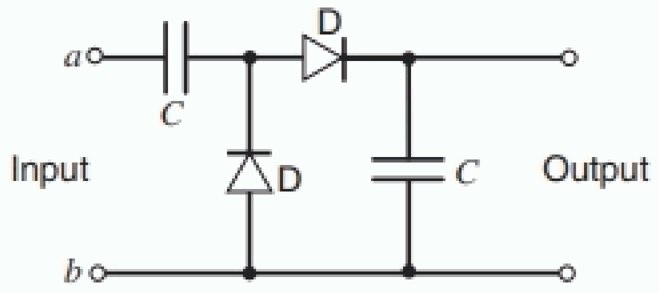


FIGURE 30-33. Problem 7.

8. A capacitor has square plates, each of side a, making an angle θ with each other as shown in Fig. 30-34. Show that for small θ the capacitance is given by

$$C = \frac{\epsilon_0 a^2}{d}\left(1 - \frac{a\theta}{2d}\right).$$

(Hint: The capacitor may be divided into differential strips that are effectively in parallel.)

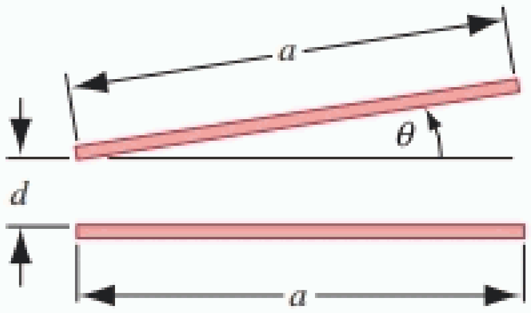


FIGURE 30-34. Problem 8.

9. In Fig. 30-35, a battery supplies a potential difference ΔV of 12 V. (*a*) Find the charge on each capacitor when switch S_1 is closed and (*b*) when (later) switch S_2 is also closed. Take $C_1 = 1.0\ \mu F$, $C_2 = 2.0\ \mu F$, $C_3 = 3.0\ \mu F$, and $C_4 = 4.0\ \mu F$.

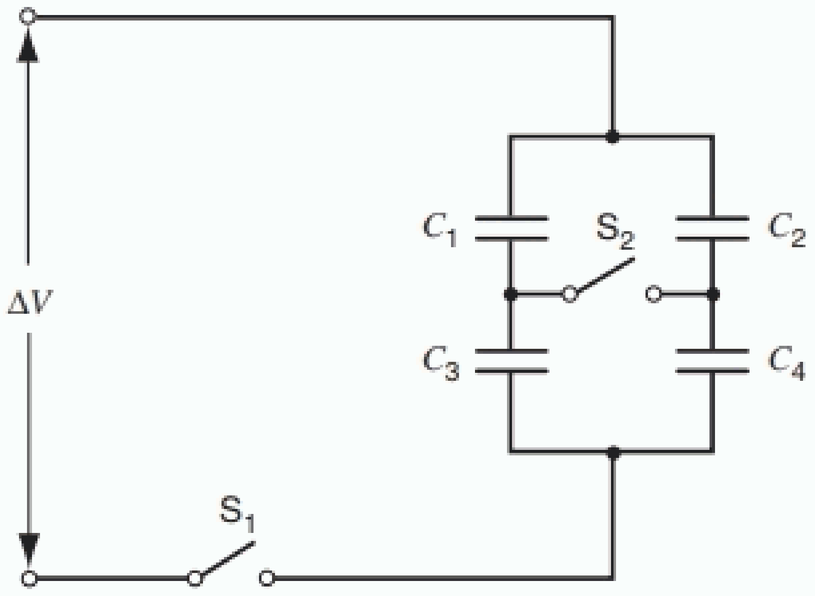


FIGURE 30-35. Problem 9.

10. Find the equivalent capacitance between points x and y in Fig. 30-36. Assume that $C_2 = 10\ \mu F$ and that the other capacitors are all 4.0 μF. (Hint: Apply a potential difference ΔV between x and y and write down all the relationships that involve the charges and potential differences for the separate capacitors.)

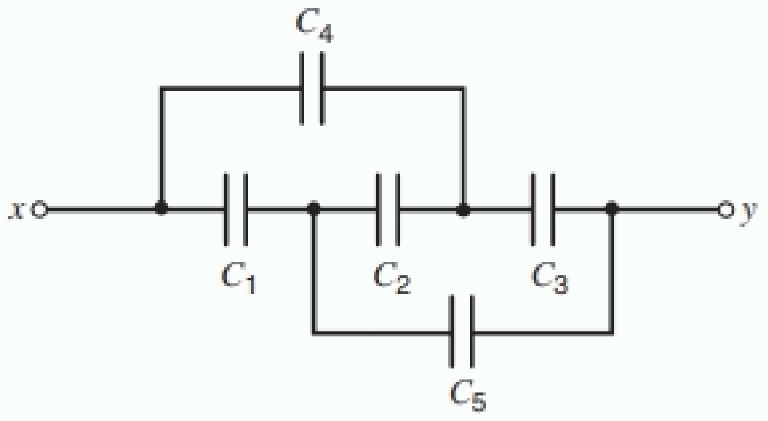


FIGURE 30-36. Problem 10.

11. One capacitor is charged until its stored energy is 4.0 J, and the charging battery is then removed. A second uncharged capacitor is then connected to it in parallel. (*a*) If the charge distributes equally, what is now the total energy stored in the electric fields? (*b*) Where did the excess energy go?

12. A fluid with resistivity 9.40 $\Omega \cdot m$ seeps into the space between the plates of a 110-pF parallel-plate, air-filled capacitor. When the space is completely filled, what is the resistance between the plates?

13. (*a*) Calculate the energy density of the electric field at a distance r from an electron (presumed to be a particle) at rest. (*b*) Assume now that the electron is not a point but a sphere of radius R over whose surface the electron charge is uniformly distributed. Determine the energy associated with the external electric field in vacuum of the electron as a function of R. (*c*) If you now associate this energy with the mass of the electron, you can, using $E_0 = mc^2$, calculate a value for R. Evaluate this radius numerically; it is often called the classical radius of the electron.

14. Show that the plates of a parallel-plate capacitor attract each other with a force given by

$$F = \frac{q^2}{2\epsilon_0 A}.$$

Prove this by calculating the work necessary to increase the plate separation from x to $x + dx$, the charge q remaining constant.

15. Using the result of Problem 14, show that the force per unit area (the *electrostatic stress*) acting on either capacitor plate is given by $\frac{1}{2}\epsilon_0 E^2$. Actually, this result is true, in general, for a conductor of any shape with an electric field $\vec{\mathbf{E}}$ at its surface.

16. A soap bubble of radius R_0 is slowly given a charge q. Because of mutual repulsion of the surface charges, the radius increases slightly to R. The air pressure inside the bubble drops, because of the expansion, to $p(V_0/V)$, where p is the atmospheric pressure, V_0 is the initial volume, and V is the final volume. Show that

$$q^2 = 32\pi^2\epsilon_0 pR(R^3 - R_0^3).$$

[Hint: Consider the forces acting on a small area of the charged bubble, ignoring surface tension. These are due to (i) gas pressure, (ii) atmospheric pressure, (iii) electrostatic stress. See Problem 15.]

17. A cylindrical ionization chamber has a central wire anode of radius 0.180 mm and a coaxial cylindrical cathode of radius 11.0 mm. It is filled with a gas whose dielectric strength is 2.20 MV/m. Find the largest potential difference that should be applied between anode and cathode if the gas is to avoid electric breakdown before radiation penetrates the chamber window.

18. A parallel-plate capacitor is filled with two dielectrics as in Fig. 30-37. Show that the capacitance is given by

$$C = \frac{\epsilon_0 A}{d}\left(\frac{\kappa_{e1} + \kappa_{e2}}{2}\right).$$

Check this formula for all the limiting cases that you can think of. (Hint: Can you justify regarding this arrangement as two capacitors in parallel?)

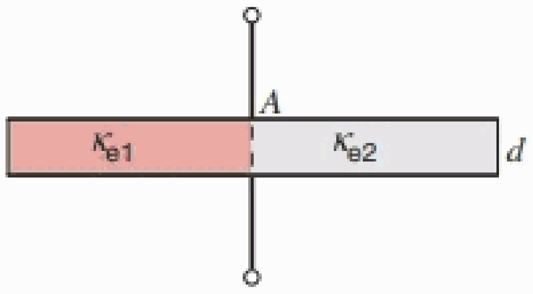


FIGURE 30-37. Problem 18.

19. A parallel-plate capacitor is filled with two dielectrics as in Fig. 30-38. Show that the capacitance is given by

$$C = \frac{2\epsilon_0 A}{d}\left(\frac{\kappa_{e1}\kappa_{e2}}{\kappa_{e1} + \kappa_{e2}}\right).$$

Check this formula for all the limiting cases that you can think of. (Hint: Can you justify regarding this arrangement as two capacitors in series?)

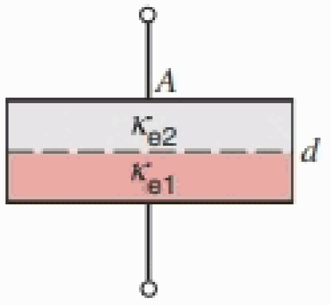


FIGURE 30-38. Problem 19.

20. What is the capacitance of the capacitor in Fig. 30-39?

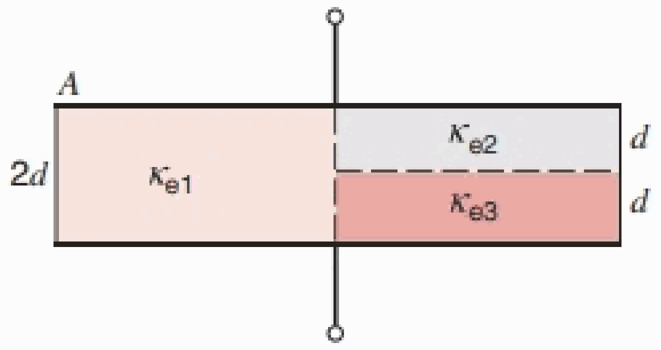


FIGURE 30-39. Problem 20.

21. A parallel-plate capacitor has plates of area A and separation d and is charged to a potential difference ΔV. The charging battery is then disconnected and the plates are pulled apart until their separation is $2d$. Derive expressions in terms of A, d, and ΔV for (*a*) the new potential difference, (*b*) the initial and the final stored energy, and (*c*) the work required to separate the plates.

22. In the capacitor of Sample Problem 30-9 (Fig. 30-15), (*a*) what fraction of the energy is stored in the air gaps? (*b*) What fraction is stored in the slab?

23. A parallel-plate capacitor has plates of area 0.118 m^2 and a separation of 1.22 cm. A battery charges the plates to a potential difference of 120 V and is then disconnected. A dielectric slab of thickness 4.30 mm and dielectric constant 4.80 is then placed symmetrically between the plates. (*a*) Find the capacitance before the slab is inserted. (*b*) What is the capacitance with the slab in place? (*c*) What is the free charge q before and after the slab is inserted? (*d*) Determine the electric field in the space between the plates and dielectric. (*e*) What is the electric field in the dielectric? (*f*) With the slab in place, what is the potential difference across the plates? (*g*) How much external work is involved in the process of inserting the slab?

24. A dielectric slab of thickness b is inserted between the plates of a parallel-plate capacitor of plate separation d. Show that the capacitance is given by

$$C = \frac{\kappa_e \epsilon_0 A}{\kappa_e d - b(\kappa_e - 1)}.$$

(Hint: Derive the formula following the pattern of Sample Problem 30-9.) Does this formula predict the correct numerical result of that sample problem? Verify that the formula gives reasonable results for the special cases of $b = 0$, $\kappa_e = 1$, and $b = d$.

COMPUTER PROBLEMS

1. Use the results of Exercise 19 to find the equivalent capacitance between any two points (such as *A* and *B*) separated by a single capacitor for the infinite capacitor array in Fig. 30-26. This problem is easiest done by iteration, and can be programmed and solved on a spreadsheet in less than a minute!

2. Repeat Computer Problem 1 for a torus instead of an infinite sheet. Start with a square grid of 10 × 10 capacitors. Join two opposite edges to make a cylinder, then join the two ends to make a torus (a "donut" shape). By how much does the answer change if the original grid is doubled in size?

CHAPTER 31

DC CIRCUITS

In Chapter 29 we discussed some general properties of current and resistance. In this chapter we begin to study the behavior of specific electric circuits that include resistive elements, which may be individual resistors or the internal resistances of circuit elements such as batteries or wires.

We confine our study in this chapter to direct current *(DC) circuits, in which the direction of the current does not change with time. In DC circuits that contain only batteries and resistors, the magnitude of the current does not vary with time, whereas in DC circuits containing capacitors, the magnitude of the current may be time dependent. Alternating current (AC) circuits, in which the current changes direction periodically, are considered in Chapter 37.*

31-1 ELECTRIC CURRENT

In Section 29-3, we discussed the flow of electric charge through conductors. The electric current i is the net amount of charge per unit time passing through an element of surface area at any particular location in the conductor. More often, we are concerned with the total current passing through a circuit, in which case the area is the entire cross section of the wires that connect the parts of the circuit.

Figure 31-1 illustrates the general problem we analyze in this chapter. A battery is connected to a "device." The device may be a single circuit element, such as a resistor or a capacitor, or it may be a combination of circuit elements. The battery maintains the upper terminal at a potential V_+ and the lower terminal at a potential V_-. For an ideal battery, the potential difference $V_+ - V_-$ between its terminals is independent of the amount of current that it is providing to the circuit. As we discuss later in this chapter, for real batteries the potential difference *does* depend on the current.

In the electrostatic case, in which conductors are equipotentials, the potential V_+ at the positive terminal of the battery would characterize the entire wire connecting the upper end of the device to the battery. In this case, the potential difference $V_+ - V_-$ between the battery terminals would also appear between the upper and lower terminals of the device. When currents are flowing in the wires, the conclusions of electrostatics are no longer valid; in particular, we know from our discussion in Section 29-4 (see Eq. 29-12) that when a current i flows in a conductor there is a potential difference $\Delta V = iR$ across the conductor. However, the

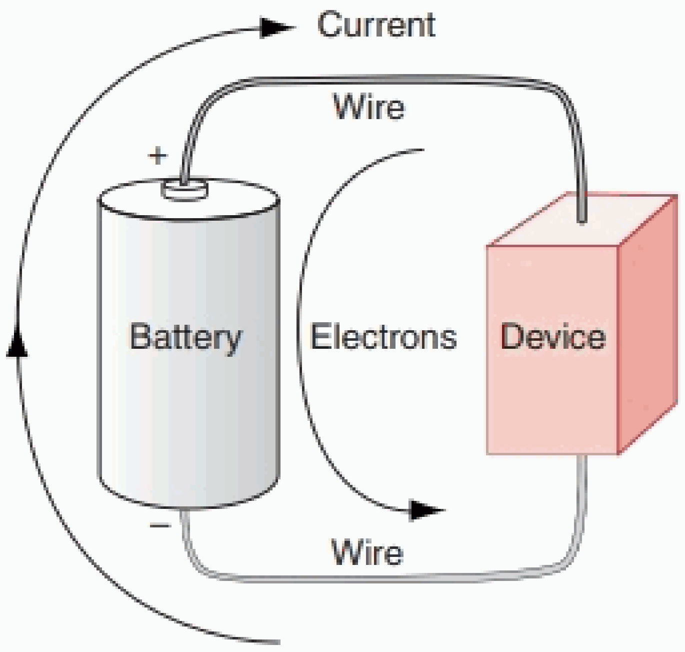


FIGURE 31-1. A battery is connected to an electrical device by two wires. The direction of the current is opposite to the direction of motion of the electrons.

resistance of the wires is usually very small compared with the resistance of the device in our circuit, so we are usually justified in neglecting the effect of the wires; in particular, we assume that there is no potential drop in the wires, and in this case the full voltage difference of the battery terminals does appear across the terminals of the device.

The battery can be considered a "pump" for charge, as if it were bringing positive charge through the battery from the negative terminal to the positive one. In actuality, it is usually the motion of the negatively charged electrons that is responsible for the current flow. Another way of interpreting the flow of charge in the circuit is to regard the positive charge as "falling" through the device from a region of high potential (the part of the device connected to the positive terminal of the battery) to a region of lower potential (the part of the device connected to the negative terminal of the battery).

The function of the battery in the circuit is to maintain the potential difference that enables the flow of charge. The battery is *not* a source of electrons. Electrons pass through the battery and have their energy raised as the move inside the battery from the positive to the negative terminal. When we say that a battery is "drained," we do not mean that it has "used up" its supply of electrons; instead, we mean that we have exhausted the source of energy (often a chemical reaction) that was responsible for raising the energies of the electrons. Note in Fig. 31-1 that the electrons move throughout the entire circuit; they do not "come from" the battery.

Although in metals the charge carriers are electrons, in electrolytes or in gaseous conductors (plasmas) they may also be positive or negative ions, or both. We need a convention for labeling the direction of current because charges of opposite sign move in opposite directions in a given field. A positive charge moving in one direction is equivalent in nearly all external effects to a negative charge moving in the opposite direction. Hence, for simplicity and algebraic consistency, we adopt the following convention:

> *The direction of current is the direction that positive charges would move, even if the actual charge carriers are negative.*

If the charge carriers are negative, they simply move opposite to the direction of the current arrow (as in Fig. 31-1).

Under most circumstances, we analyze electric circuits based on an assumed direction for the current, without taking into account whether the actual charge carriers are positive or negative. In rare cases (see, for example, the Hall effect in Section 32-4) we must take into account the sign of the charge carriers.

Conservation of Charge

When we first connect the battery to the device, the circuit behaves in an irregular manner. The situation is similar to what occurs when you first turn on a garden hose connected to a sprinkler. At first the water gushes through the hose, creating whirlpools and eddies. When it reaches the sprinkler it may at first by chance emerge from some of the holes and not from others. After a few seconds a steady flow is established, and the water flows past any point at a constant rate. In electrical circuits, we usually ignore this initial behavior (called the *transient behavior*) and consider only the steady situation, which is reached very rapidly (within nanoseconds).

We assume that, under steady conditions, charge does not collect at or drain away from any point in our idealized wire. In the language of fluid flow, there are no sources or sinks of charge in the wire. When we made this assumption in our study of incompressible fluids, we concluded that the rate at which the fluid flows past any cross section of a pipe is the same even if the cross section varies. The fluid flows faster where the pipe is smaller and slower where it is larger, but the volume rate of flow, measured perhaps in liters/second, remains constant. In the same way, *the electric current i is the same for all cross sections of a conductor, even though the cross-sectional area may be different at different points.* The current density $\vec{\mathbf{j}}$ (current per unit area) will change as the cross-sectional area changes, but the current i remains the same.

Figure 31-2*a* shows the circuit of Fig. 31-1 in simpler notation. The device is indicated merely by a box. Note that the current going into the device is the same as the current coming out of the device. This is an example of conservation of charge; no net charge is retained by the device—for every electron that enters at one end, one electron leaves the other end.

Figure 31-2*b* shows another circuit in which the current travels in succession through three devices, labeled A, B, and C. The current i_A in device A is exactly equal to the current i_B in device B and also to the current i_C in device C; that is, $i_A = i_B = i_C$. No current is "used up" in traveling through any circuit element. Figure 31-2*b* is an example of circuit elements connected in *series,* in which the same current must travel in succession through each element of the circuit.

Figure 31-2*c* shows the current in a different combination of circuit elements. Here the current must divide when it reaches point a in the circuit, with an amount i_A passing through device A and an amount i_B passing through device B. (The relative amounts in A and B, which are not important for this discussion, depend on the properties of A and B.) At point b the currents must recombine. Because no charge gets trapped at point a, the current entering point a must be exactly the same as the current leaving it, or $i = i_A + i_B$. Similarly, the current entering point b must be the same as the current leaving point b, or $i_A + i_B = i$. This is often called the *junction rule* for analyzing circuits:

> *At any junction in an electric circuit, the total current entering the junction must be equal to the total current leaving the junction.*

In this rule the term "junction" means a point in a circuit where several wire segments meet, such as points a or b in Fig. 31-2*c*. The junction rule (sometimes known as *Kirch-*

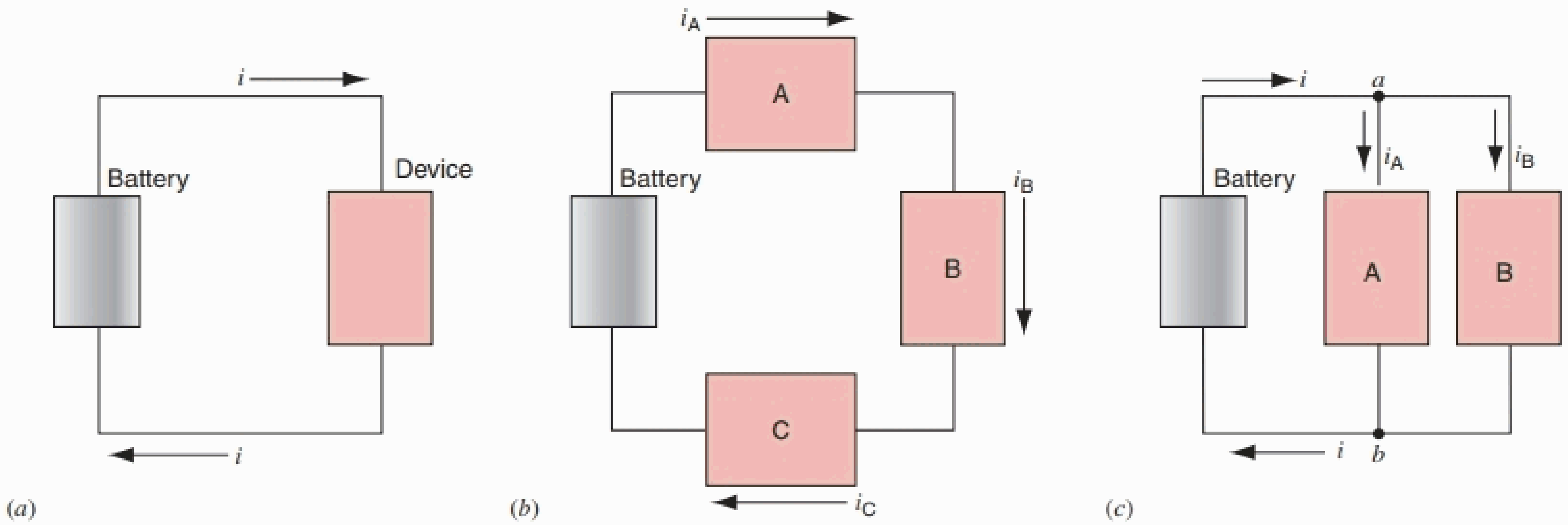


FIGURE 31-2. (*a*) The circuit of Fig. 31-1 in symbolic notation. (*b*) The same current flows in succession through devices A, B, and C. (*c*) The current divides at junction *a* and recombines at junction *b*.

hoff's first law) is really a statement about the conservation of electric charge.

Figure 31-2*c* is an example of a *parallel* connection of circuit elements. A characteristic of a parallel connection is that the current must divide to pass separately through individual elements and then recombine.

31-2 ELECTROMOTIVE FORCE

An external energy source is required by most electrical circuits to move charge through the circuit. The circuit therefore must include a device that maintains a potential difference between two points in the circuit, just as a circulating fluid requires an analogous device (a pump) that maintains a *pressure* difference between two points.

Any device that performs this task in an electrical circuit is called a source (or a seat) of *electromotive force* (symbol $\mathcal{E}$; abbreviation emf, which is pronounced "ee-em-eff"). It is sometimes useful to consider a source of emf as a mechanism that creates a "hill" of potential and moves charge "uphill," from which the charge flows "downhill" through the rest of the circuit. A common source of emf is the ordinary battery; another is the electric generator found in power plants. Solar cells are sources of emf used both in spacecraft and in pocket calculators. Other less commonly found sources of emf are fuel cells (used to power the space shuttle) and thermopiles. Biological systems, including the human heart, also function as sources of emf.

Figure 31-3*a* shows a source of emf connected to an electronic device, which might be a resistor, capacitor, or other circuit element. The emf is represented in the circuit by an arrow that is placed next to the source and points in the direction in which the emf, acting alone, would cause a positive charge carrier to move in the external circuit. We draw a small circle on the tail of the emf arrow so that it will not be mistaken for a current arrow. In the external circuit, positive charge carriers would be driven in the direction shown by the current arrows marked *i*. In other words, the source of emf sets up a clockwise current in the circuit of Fig. 31-3*a*.

The most common source of emf that we will use is the ordinary battery. As shown in Fig. 31-3*a*, a battery is represented in a circuit by two parallel lines of different length. The longer line always indicates the positive terminal of the battery. The source of emf (the battery) maintains its upper terminal at a high potential V_+ and its lower terminal at a low potential V_-. Common batteries (size AAA, AA, C, or D) that we use for flashlights or portable CD players have an emf of 1.5 volts. As we will see in the next section, the battery emf is equal to the terminal potential difference $V_+ - V_-$ *only* if there is no current in the circuit or if the battery has a negligible internal resistance.

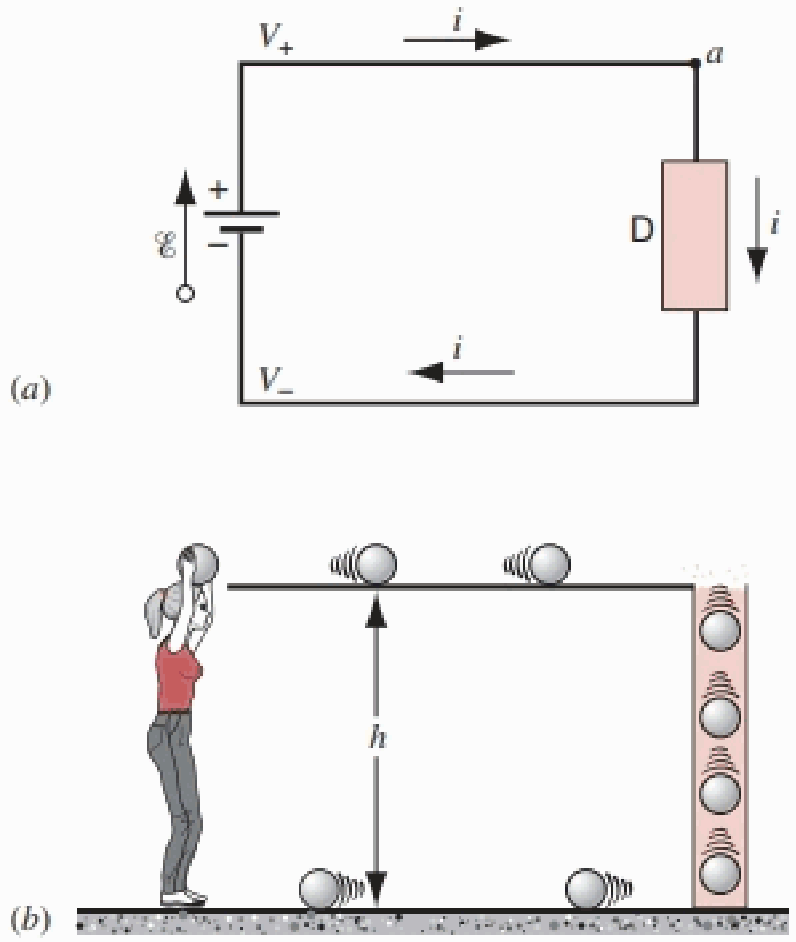


FIGURE 31-3. (*a*) A simple electric circuit, in which the emf $\mathcal{E}$ (a battery) does work on the charge carriers and maintains a steady current through device D. (*b*) A gravitational analogue, in which work done by the person maintains a steady flow of bowling balls through the viscous medium.

A source of emf must be able to do work on charge carriers that enter it. The source acts to move positive charges from a point of low potential (the negative terminal) through its interior to a point of high potential (the positive terminal). The charges then move through the external circuit, in the process dissipating the energy provided to them by the source of emf. Eventually, they return to the negative terminal, from which the emf raises them to the positive terminal again, and the cycle continues. (Note that, in accordance with the usual convention, we analyze the circuit as if positive charge were flowing. The actual motion of the electrons is in the opposite direction.)

When a steady current has been established in the circuit of Fig. 31-3*a*, a charge dq passes through *any* cross section of the circuit in time dt. In particular, this charge enters the source of emf $\mathcal{E}$ at its low-potential end and leaves at its high-potential end. The source must do an amount of work dW on the (positive) charge carriers to force them to go to the point of higher potential. The emf $\mathcal{E}$ of the source is defined as the work per unit charge, or

$$\mathcal{E} = dW/dq. \tag{31-1}$$

The unit of emf is the joule/coulomb, which is the same as the volt. Note from Eq. 31-1 that the electromotive force is not actually a force; that is, we do not measure it in newtons. The name originates from the early history of the subject.

The source of emf provides energy to the circuit. Its energy might be obtained from a variety of processes: chemical (as in a battery or a fuel cell), mechanical (a generator), thermal (a thermopile), or radiant (a solar cell). The current in the circuit of Fig. 31-3*a* transfers energy from the source of emf to the device D. If the device is another battery that is being charged by the source of emf, then the energy transferred appears as chemical energy newly stored in the battery. If the device is a resistor, the transferred energy appears as internal energy (observed perhaps as an increase in temperature) and can then be transferred to the environment as heat. If the device is a capacitor, the energy transferred is stored as potential energy in its electric field. In each of these cases, conservation of energy requires that the amount of energy lost by the battery must equal the amount of energy transferred to, dissipated by, or stored in the device D.

Figure 31-3*b* shows a gravitational analogue of Fig. 31-3*a*. In the lower figure the person, in lifting the bowling balls from the floor to the shelf, does work on them. The balls roll slowly and uniformly along the shelf, dropping from the right end into a cylinder full of viscous oil. They sink to the bottom at an essentially constant speed, are removed by a trapdoor mechanism not shown, and roll back along the floor to the left. The energy put into the system by the person appears eventually as internal energy in the viscous fluid, resulting in a temperature rise. The energy supplied by the person comes from her store of internal (chemical) energy. The circulation of charges in Fig. 31-3*a* stops eventually if the source of emf runs out of energy; the circulation of bowling balls in Fig. 31-3*b* stops eventually if the person runs out of energy.

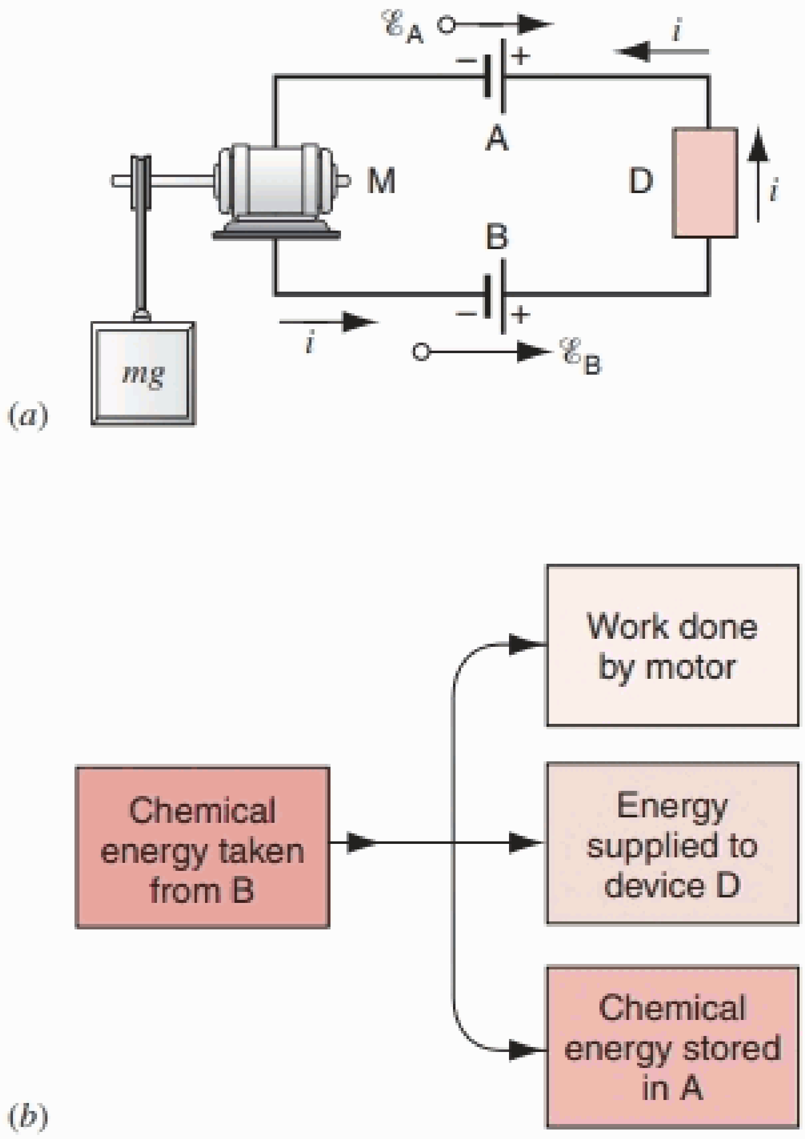

$\mathcal{E}_A$
i
A
M
D
i
B
i
$\mathcal{E}_B$
mg
(*a*)
Work done by motor
Chemical energy taken from B
Energy supplied to device D
Chemical energy stored in A
(*b*)

FIGURE 31-4. (*a*) $\mathcal{E}_B > \mathcal{E}_A$, so that battery B determines the direction of the current in this single-loop circuit. (*b*) Energy transfers in this circuit.

Figure 31-4*a* shows a circuit containing two sources of emf (batteries) A and B, a device D, and an ideal electric motor M employed in lifting a weight. The batteries are connected so that they tend to send charges around the circuit in opposite directions; the actual direction of the current is determined by battery B, which has the larger emf. Figure 31-4*b* shows the energy transfers in this circuit. The chemical energy in battery B is steadily depleted, the energy appearing in the three forms shown on the right. Battery A is being charged while battery B is being discharged.

Note that the circuit can transfer energy *from* a source of emf or *into* a source of emf. In the ideal case, the energy transfer process associated with a source of emf is *reversible* in the thermodynamic sense (see Section 24-1). A battery can either be *charged* (meaning an external source adds to the battery's supply of energy, *not* that we are forcing more charge into the battery) or *discharged* (meaning we take energy from the battery). Similarly, a generator can be driven mechanically to produce electrical energy, or it can use electrical energy to produce mechanical motion, as in a motor.

31-3 ANALYSIS OF CIRCUITS

The simplest electrical circuit consists of one source of emf (such as a battery) and one circuit device (such as a resistor). Examples of this kind of circuit might include a flashlight or an electric heater. Figure 31-5 shows a circuit con-

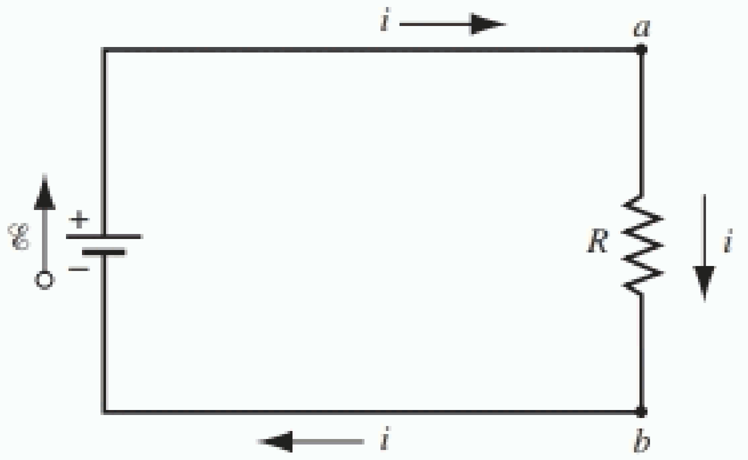


FIGURE 31-5. A single-loop circuit. The current is the same everywhere in the circuit. The potential *increases* from − to + across the battery, and it *decreases* from *a* to *b* (in the direction of the current) across the resistor.

sisting of a single ideal battery and a resistor R. The symbolic circuit notation for a resistor is .

Often our goal in analyzing circuits is to determine the magnitude and direction of the current, given the emfs and resistors in the circuit. We will analyze this circuit by considering the potential differences across each circuit element. Later in this chapter we consider another method based on the energy supplied or dissipated by each circuit element.

The first step in analyzing the circuit is to guess a direction for the current. Usually we try to make the best guess we can; if we choose the wrong direction the current will come out to be negative, indicating that our original guess for the direction was wrong, but the magnitude we calculate will still be correct. In the circuit of Fig. 31-5, we expect that the current is clockwise, determined by the emf of the battery. The battery maintains point a at a higher potential than point b, and so positive charges in the circuit would "fall" through the resistor from a to b before returning to the battery to be pumped back up to the higher potential of a.

When we analyze the circuit using the method of potential differences, we travel once around the loop and keep track of the differences in potential across each circuit element. It does not matter which way we travel around the loop in making this analysis. Let us try going around clockwise (in the same direction as the current), starting at point a.

Let ΔV_R be the potential difference across the resistor. That is, the potential at a is greater than the potential at b by an amount ΔV_R, which is equal to iR. How do we know that V_a is greater than V_b? We do not yet know this for certain, but it is consistent with the direction we have guessed for the current through R. If the current flows from a to b, then the positive charge carriers are "falling" through the resistor from the higher potential at a to the lower potential at b. If our initial guess for the direction of the current is wrong, our solution will show that both i and ΔV_R are negative.

We are now ready to analyze the circuit. Our procedure is to start at any point, travel once around the circuit adding all the potential differences, and then return to the starting point where we must find the same potential at which we started. This procedure can be summarized as follows:

> *The algebraic sum of all differences in potential around a complete circuit loop must be zero.*

This rule is known as the *loop rule* (and is sometimes referred to as *Kirchhoff's second law*). Ultimately it is a statement about the conservation of energy. An analogy would be taking a walk in a hilly terrain; you may travel up or down as you walk, but if you keep track of all your changes in gravitational potential energy you will find that the total change is zero when you return to your starting point.

We now examine the changes in potential as we go once around the circuit of Fig. 31-5, starting at point a where the potential is V_a. Proceeding clockwise from a through the resistor, the potential drops by $\Delta V_R = iR$, so the potential at b is $V_b = V_a - iR$. Continuing clockwise around the circuit, we next travel through the battery from the negative terminal to the positive terminal, and thus the potential increases by the battery emf $\mathcal{E}$. This now brings us back to point a and the potential V_a. Since the starting and ending potentials at point a must be equal (potential being a path-independent quantity), we therefore have $V_a = V_a - iR + \mathcal{E}$. Equivalently, we can find this result by applying the loop rule directly, adding up the differences in potential, and setting the resulting sum equal to zero. Again starting at a and proceeding clockwise, we first encounter a negative potential difference of $-iR$ and then a positive potential difference of $+\mathcal{E}$. Setting the sum of these potential differences to zero, we obtain

$$-iR + \mathcal{E} = 0$$

or

$$i = \frac{\mathcal{E}}{R}. \tag{31-2}$$

We have now found the current in this circuit, which completes our analysis.

Let us now consider a slightly more complicated single-loop circuit, shown in Fig. 31-6. This circuit has a single battery but two resistors. We can once again guess that the current is clockwise. Let us travel around the circuit in a

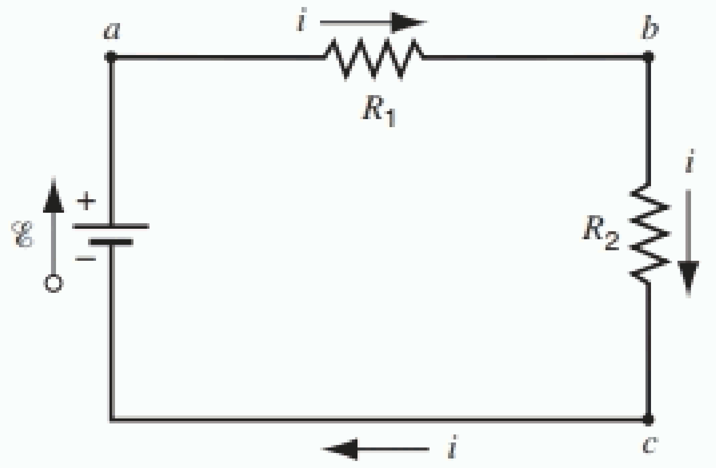


FIGURE 31-6. A single-loop circuit with two resistors. The current is the same everywhere; the potential decreases from *a* to *b* and also from *b* to *c*, in the direction of the current.

counterclockwise direction this time. Starting at point *a*, we first go through the battery and find a potential difference of $-\mathscr{E}$. Next we go through R_2 in a direction opposite to the current, so the potential *increases* and the potential difference is $+iR_2$. Similarly, the potential difference when we go through R_1 is $+iR_1$, after which we are back at our starting point. According to the loop rule, the total of these potential differences is zero:

$$-\mathscr{E} + iR_2 + iR_1 = 0$$

or

$$i = \frac{\mathscr{E}}{R_1 + R_2}. \tag{31-3}$$

Note that Eq. 31-3 reduces to Eq. 31-2 if either $R_1 = 0$ or $R_2 = 0$.

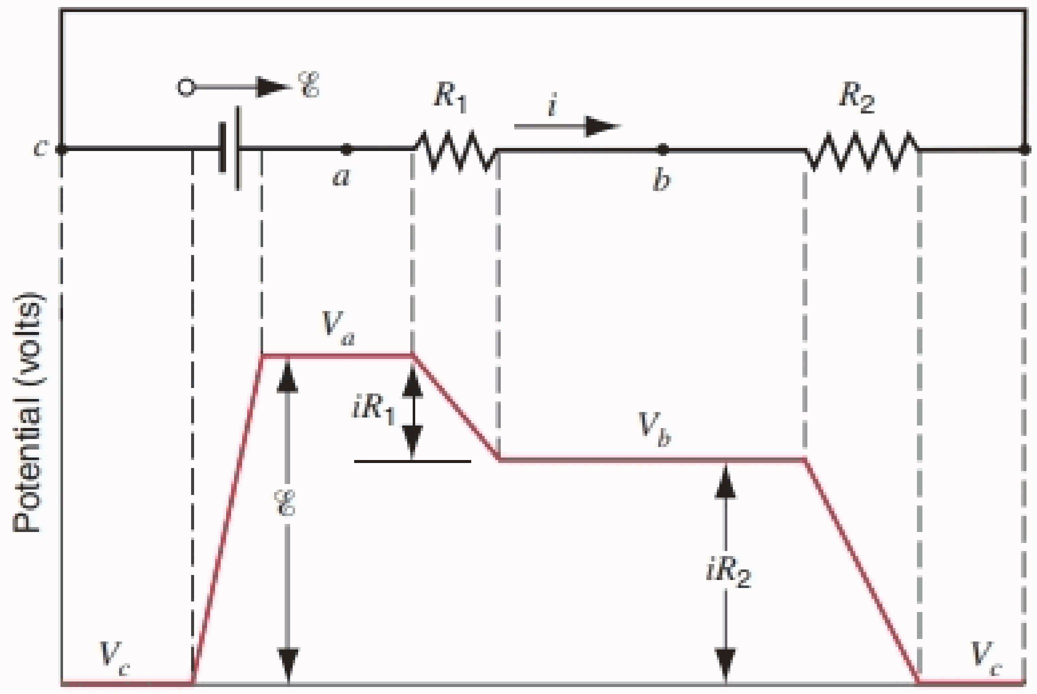


FIGURE 31-7. The circuit of Fig. 31-6 is drawn with its components along a straight line at the top. The potential difference across each of the elements is shown.

Potential Differences in a Circuit

We often want to find the potential difference between two points in a circuit. In Fig. 31-6, for example, how does the potential difference ΔV_{ab} $(= V_a - V_b)$ between points *b* and *a* depend on the circuit elements $\mathscr{E}$, R_1 and R_2? To find their relationship, let us start at point *b* and move counterclockwise to point *a*, passing through resistor R_1. If V_a and V_b are the potentials at *a* and *b*, respectively, we have

$$V_b + iR_1 = V_a,$$

because we experience an increase in potential in traveling through a resistor in the direction opposite to the current. We rewrite this relation in terms of ΔV_{ab}, the potential difference between *a* and *b*, as

$$\Delta V_{ab} = V_a - V_b = +iR_1,$$

which tells us that ΔV_{ab} has magnitude iR_1 and that point *a* is at a higher potential than point *b*. Combining this last equation with Eq. 31-3 yields

$$\Delta V_{ab} = \mathscr{E}\frac{R_1}{R_1 + R_2}. \tag{31-4}$$

In summary, to find the potential difference between any two points in a circuit, start at one point, travel through the circuit to the other, and add algebraically the changes in potential encountered. This algebraic sum is the potential difference between the points. This procedure is similar to that for finding the current in a closed loop, except that here the potential differences are added over part of a loop and not over the whole loop.

You can travel *any* path through the circuit between the two points, and you get the same value of the potential difference, because *path independence* is an essential part of our concept of potential. The potential difference between any two points can have only one value; we must obtain the same result for all paths that connect those points. (Similarly, if we consider two points on the side of a hill, the measured difference in gravitational potential between them is the same no matter what path is followed in going from one to the other.) In Fig. 31-6 let us recalculate ΔV_{ab}, using a path starting at *a* and going counterclockwise through the source of emf to *b*. We have

$$V_a - \mathscr{E} + iR_2 = V_b$$

or

$$\Delta V_{ab} = V_a - V_b = +\mathscr{E} - iR_2.$$

Combining this result with Eq. 31-3 leads to Eq. 31-4.

Using similar methods, we can show that

$$\Delta V_{bc} = \mathscr{E}\frac{R_2}{R_1 + R_2}. \tag{31-5}$$

Note that, as we should expect, $\Delta V_{ab} + \Delta V_{bc} = \mathscr{E}$. The combination of resistors in the circuit of Fig. 31-6 is called a *voltage divider.* In effect, it divides the voltage difference of the battery into two pieces in proportion to the sizes of the two resistors.

Another way of illustrating the potential differences in this circuit is shown in Fig. 31-7. For convenience we have started at point *c* and traveled clockwise around the circuit. Here you can clearly see how the battery emf is "divided" into the potential differences across the two resistors.

Internal Resistance of a Source of EMF

In contrast to the ideal batteries we have been considering so far, real batteries have an internal resistance. This resistance is characteristic of the materials of which the battery is made. Because it is an inherent part of the battery, it cannot be removed; we would usually like to do so, for internal resistance has undesirable effects such as reducing the terminal voltage of the battery and limiting the current that can flow in the circuit.

Figure 31-8 shows the simple loop circuit of Fig. 31-5 with the internal resistance *r* of the battery taken into account. Even though they are part of the same device, we

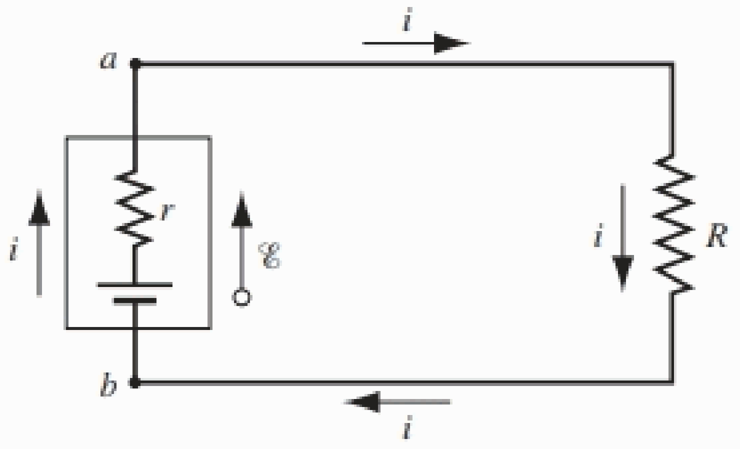


FIGURE 31-8. A battery is represented as a device containing a source of emf ℰ and an internal resistance r.

show the source of emf and the internal resistance as separate elements.

The circuit of Fig. 31-8 is identical to the circuit of Fig. 31-6, and we can find the current by merely adapting Eq. 31-3 to the circuit elements of Fig. 31-8:

$$i = \frac{\mathscr{E}}{R + r}. \tag{31-6}$$

The internal resistance *reduces* the current that the emf can supply to the external circuit.

The potential difference between the battery terminals is $\Delta V_{ab} = V_a - V_b = \mathscr{E} - iR$; using Eq. 31-6 we obtain

$$\Delta V_{ab} = \mathscr{E}\frac{R}{R + r}. \tag{31-7}$$

From this expression we see that the potential difference between the battery terminals is not a constant, but now depends on the resistance R of the external circuit. As we make R smaller, thereby increasing the current, the potential difference between the battery terminals decreases. A 1.5-V battery has a terminal voltage difference of 1.5 V only when there is no current flowing through the battery. When the battery is connected to a circuit device such as a radio, the voltage difference between the terminals will be less than 1.5 V.

We see from Eq. 31-7 that ΔV_{ab} is equal to ℰ only if either the battery has no internal resistance ($r = 0$) or the external circuit is open ($R = \infty$).

SAMPLE PROBLEM 31-1. What is the current in the circuit of Fig. 31-9*a*? The emfs and the resistors have the following values: $\mathscr{E}_1 = 2.1$ V, $\mathscr{E}_2 = 4.4$ V, $r_1 = 1.8\ \Omega$, $r_2 = 2.3\ \Omega$, $R = 5.5\ \Omega$.

Solution The two emfs are connected so that they oppose each other but $\mathscr{E}_2$, because it is larger than $\mathscr{E}_1$, controls the direction of the current in the circuit, which is counterclockwise. The loop rule, applied clockwise from point a, yields

$$-\mathscr{E}_2 + ir_2 + iR + ir_1 + \mathscr{E}_1 = 0.$$

Check that this same equation results by going around counterclockwise or by starting at some point other than a. Also, compare this equation term by term with Fig. 31-9*b*, which shows the potential changes graphically.

Solving for the current i, we obtain

$$i = \frac{\mathscr{E}_2 - \mathscr{E}_1}{R + r_1 + r_2} = \frac{4.4\ \text{V} - 2.1\ \text{V}}{5.5\ \Omega + 1.8\ \Omega + 2.3\ \Omega} = 0.24\ \text{A}.$$

It is not necessary to know the direction of the current in advance. To show this, let us assume that the current in Fig. 31-9*a* is clockwise—that is, opposite to the direction of the current arrow in Fig. 31-9*a*. The loop rule would then yield (going clockwise from *a*)

$$-\mathscr{E}_2 - ir_2 - iR - ir_1 + \mathscr{E}_1 = 0$$

or

$$i = -\frac{\mathscr{E}_2 - \mathscr{E}_1}{R + r_1 + r_2}.$$

Substituting numerical values yields $i = -0.24$ A for the current. The negative sign is a signal that the current is in the opposite direction from that which we have assumed.

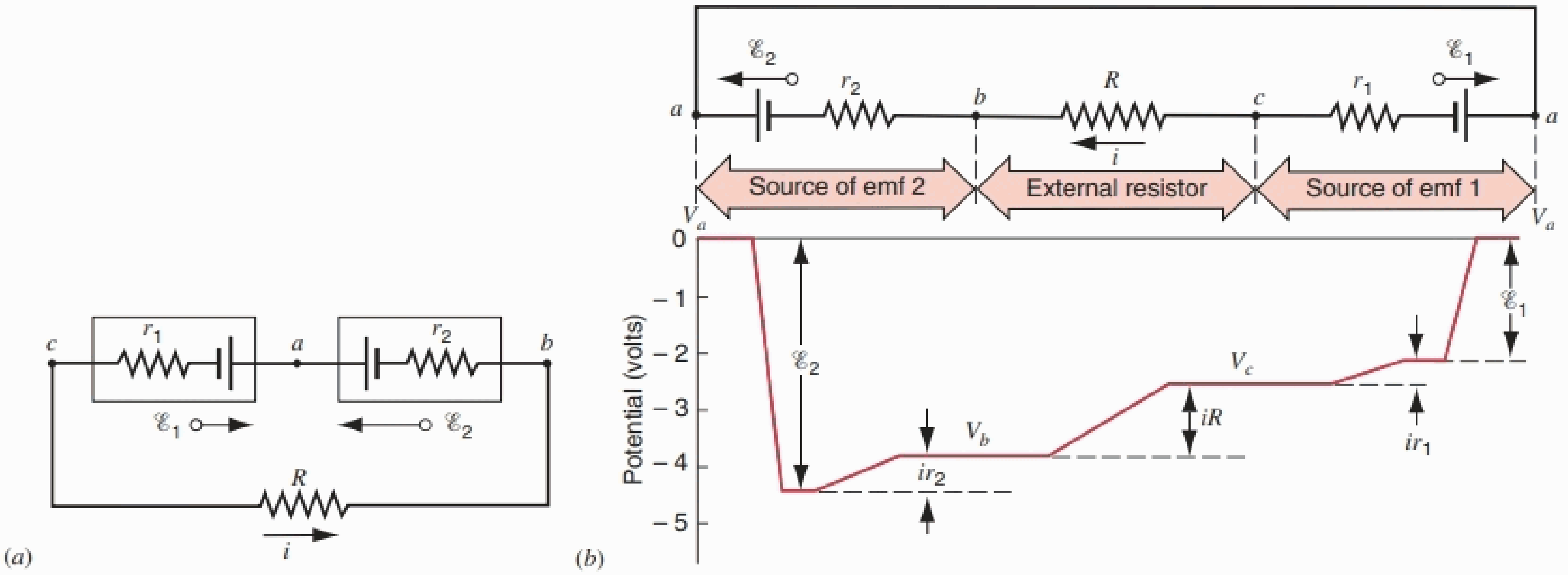


FIGURE 31-9. Sample Problems 31-1 and 31-2. (*a*) A single-loop circuit containing two sources of emf. (*b*) The changes in potential encountered in traveling clockwise around the circuit starting at point *a*.

In more complex circuits involving many loops and branches, it is often impossible to know in advance the actual directions for the currents in all parts of the circuit. However, the current directions for each branch may be chosen arbitrarily. If you get an answer with a positive sign for a particular current, you have chosen its direction correctly; if you get a negative sign, the current is opposite in direction to that chosen. In either case, the numerical value is correct.

SAMPLE PROBLEM 31-2. (*a*) What is the potential difference between points *a* and *b* in Fig. 31-9*a*? (*b*) What is the potential difference between points *a* and *c* in Fig. 31-9*a*?

Solution (*a*) This potential difference is the terminal potential difference of battery 2, which includes emf $\mathcal{E}_2$ and internal resistance r_2. Let us start at point *b* and traverse the circuit counterclockwise to point *a*, passing directly through the source of emf. We find

$$V_b - ir_2 + \mathcal{E}_2 = V_a$$

or

$$V_a - V_b = -ir_2 + \mathcal{E}_2 = -(0.24\ \text{A})(2.3\ \Omega) + 4.4\ \text{V} = +3.8\ \text{V}.$$

We see that *a* is more positive than *b* and the potential difference between them (3.8 V) is *smaller* than the emf (4.4 V); see Fig. 31-9*b*.

We can verify this result by starting at point *b* in Fig. 31-9*a* and moving through the circuit clockwise to point *a*. For this different path we find

$$V_b + iR + ir_1 + \mathcal{E}_1 = V_a$$

or

$$\begin{aligned} V_a - V_b &= iR + ir_1 + \mathcal{E}_1 \\ &= (0.24\ \text{A})(5.5\ \Omega + 1.8\ \Omega) + 2.1\ \text{V} = +3.8\ \text{V}, \end{aligned}$$

exactly as before. The potential difference between two points has the same value for all paths connecting those points.

(*b*) Note that the potential difference between *a* and *c* is the terminal potential difference of battery 1, consisting of emf $\mathcal{E}_1$ and internal resistance r_1. Let us start at *c* and traverse the circuit clockwise to point *a*. We find

$$V_c + ir_1 + \mathcal{E}_1 = V_a$$

or

$$V_a - V_c = ir_1 + \mathcal{E}_1 = (0.24\ \text{A})(1.8\ \Omega) + 2.1\ \text{V} = +2.5\ \text{V}.$$

This tells us that *a* is at a higher potential than *c*. The terminal potential difference (2.5 V) is in this case *larger* than the emf (2.1 V); see Fig. 31-9*b*. Charge is being forced through $\mathcal{E}_1$ in a direction opposite to that in which it would send charge if it were acting by itself; if $\mathcal{E}_1$ were a storage battery it would be charging at the expense of $\mathcal{E}_2$, which is being discharged.

SAMPLE PROBLEM 31-3. Figure 31-10 shows a circuit with two loops. Find the currents in the circuit. The elements have the following values: $\mathcal{E}_1 = 2.1$ V, $\mathcal{E}_2 = 6.3$ V, $R_1 = 1.7\ \Omega$, $R_2 = 3.5\ \Omega$.

Solution The first step in analyzing this circuit is to define currents and their directions in each branch. We have drawn currents i_1, i_2, and i_3 in the three branches and arbitrarily selected their directions. At point *a*, current i_3 flows in and currents i_1 and i_2 flow out. Applying the junction rule (Section 31-1), we have

$$i_3 = i_1 + i_2.$$

We now apply the loop rule to each of the two loops. Starting at *a* and going counterclockwise around the left-hand loop, we find $-i_1R_1 - \mathcal{E}_1 - i_1R_1 + \mathcal{E}_2 + i_2R_2 = 0$ or

$$2i_1R_1 - i_2R_2 = \mathcal{E}_2 - \mathcal{E}_1.$$

If we traverse the right-hand loop in a clockwise direction from point *a*, we find $+i_3R_1 - \mathcal{E}_2 + i_3R_1 + \mathcal{E}_2 + i_2R_2 = 0$ or, after substituting $i_3 = i_1 + i_2$ from the junction rule,

$$2i_1R_1 + (2R_1 + R_2)i_2 = 0.$$

We now have two equations for the two currents i_1 and i_2. We can solve these equations for these variables, obtaining, after a little algebra,

$$\begin{aligned} i_1 &= \frac{(\mathcal{E}_2 - \mathcal{E}_1)(2R_1 + R_2)}{4R_1(R_1 + R_2)} \\ &= \frac{(6.3\ \text{V} - 2.1\ \text{V})(2 \times 1.7\ \Omega + 3.5\ \Omega)}{(4)(1.7\ \Omega)(1.7\ \Omega + 3.5\ \Omega)} = 0.82\ \text{A}, \\ i_2 &= -\frac{\mathcal{E}_2 - \mathcal{E}_1}{2(R_1 + R_2)} \\ &= -\frac{6.3\ \text{V} - 2.1\ \text{V}}{(2)(1.7\ \Omega + 3.5\ \Omega)} = -0.40\ \text{A}. \end{aligned}$$

The third current can be found from applying the junction rule:

$$i_3 = i_1 + i_2 = 0.82\ \text{A} + (-0.40\ \text{A}) = 0.42\ \text{A}.$$

The signs of the currents tell us that we have guessed correctly about the directions of i_1 and i_3 but that we are wrong about the direction of i_2; it should point up—and not down—in the central branch of the circuit of Fig. 31-10.

Note that, having discovered that current i_2 is pointing in the wrong direction, we do not need to change it in Fig. 31-10. We can leave it in the figure as it is, as long as we remember to substitute a negative numerical value for i_2 in all subsequent calculations involving that current.

SAMPLE PROBLEM 31-4. What is the potential difference between points *a* and *b* in the circuit of Fig. 31-10?

Solution For the potential difference between *a* and *b*, we have, moving through branch *ab* in Fig. 31-10 and assuming the current directions shown,

$$V_a - i_2R_2 - \mathcal{E}_2 = V_b,$$

or

$$\begin{aligned} V_a - V_b &= \mathcal{E}_2 + i_2R_2 \\ &= 6.3\ \text{V} + (-0.40\ \text{A})(3.5\ \Omega) = +4.9\ \text{V}. \end{aligned}$$

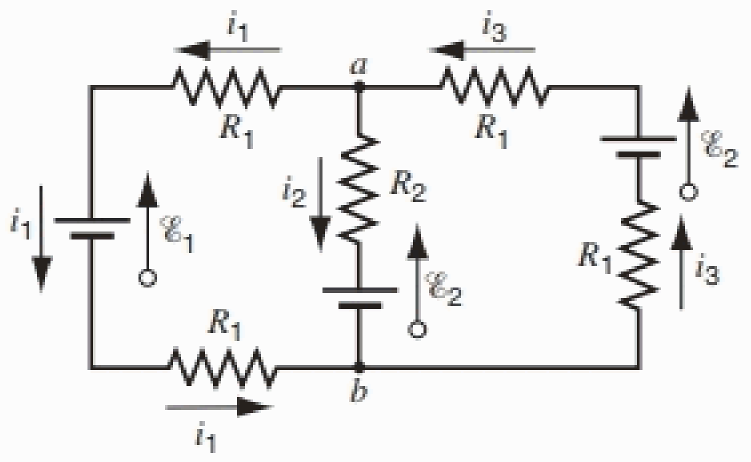


FIGURE 31-10. Sample Problems 31-3 and 31-4. A two-loop circuit.

The positive sign tells us that a is more positive in potential than b. We should expect this result from looking at the circuit diagram, because all three batteries have their positive terminals on the top side of the figure.

31-4 ELECTRIC FIELDS IN CIRCUITS*

So far we have been discussing circuits in a rather mysterious way. In Chapter 29 we considered the relationship between current and electric field in a conductor: $\vec{\mathbf{E}} = \rho\vec{\mathbf{j}}$ (Eq. 29-10), where ρ is the resistivity of the material and $\vec{\mathbf{j}}$ is the current density (current per unit cross-sectional area). The wires in our circuits are conductors, and so there must be an electric field present to establish and sustain the current. Where does this electric field come from?

It is helpful at this point to return to the analogy between current and fluid flow in a pipe. Figure 31-11 shows a closed loop similar to the circuit of Fig. 31-5. The pump is analogous to the source of emf, and the constriction in the pipe is analogous to the resistor. In the steady state, the amount of fluid per unit time passing any particular point in the circuit must be the same as that passing any other point. In the constriction, the fluid must therefore flow more rapidly (assuming the fluid to be incompressible).

The battery provides the emf to the circuit; its role is to "pump" charges from low potential to high potential. The emf is defined as the work per unit charge done by the battery. With the conventional definition of work done by a force $\vec{\mathbf{F}}$ as $W = \int \vec{\mathbf{F}} \cdot d\vec{\mathbf{s}}$, the work per unit charge W/q (the emf) must then be related to the force per unit charge $\vec{\mathbf{F}}/q$:

$$\mathcal{E} = \oint (\vec{\mathbf{F}}/q) \cdot d\vec{\mathbf{s}} \tag{31-8}$$

It is tempting to associate an electric field with the quantity $\vec{\mathbf{F}}/q$, but it is generally not correct. The force $\vec{\mathbf{F}}$ in this case is the one that acts inside the source of emf; it might be a force of mechanical, chemical, thermodynamic, or magnetic origin, but it is not necessarily associated with an electric field.

We have written Eq. 31-8 as the integral around a closed path. In this way, the emf depends only on the net effect of the source on a charge that makes a complete loop around the circuit. Conservative external fields cannot give an emf, because Eq. 31-8 vanishes for such fields.

Inside the wires, there *is* an electric field. This field must be present for charge to flow in the wires. (Recall from our discussion in Chapter 27 that the rule that $E = 0$ inside a conductor holds *only* under electrostatic conditions; when currents are present, this rule does not hold.)

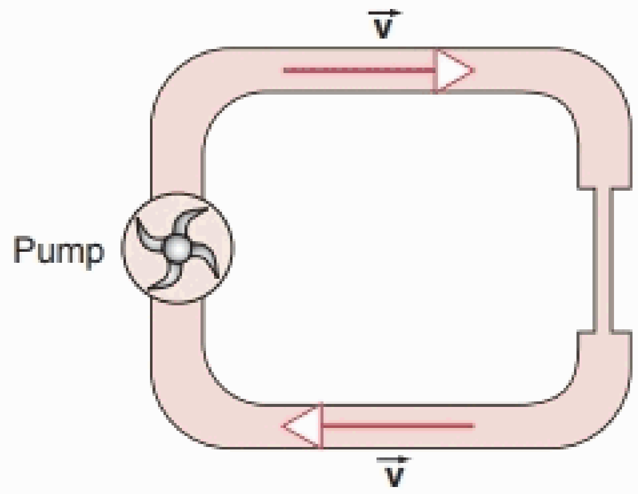


FIGURE 31-11. A fluid-mechanical analogy to the electrical circuit of Fig. 31-5. The constriction gives some resistance to the flow of fluid.

Figure 31-12 shows a representation of the electric fields in a conducting wire. How large must the electric fields be to maintain a typical current?

It is helpful to make some numerical estimates. Suppose a current $i = 1$ A flows in a wire of radius $R = 1$ mm. The current density is then

$$j = \frac{i}{A} = \frac{1\text{ A}}{\pi(1 \times 10^{-3}\text{ m})^2} \approx 3 \times 10^5\text{ A/m}^2.$$

For copper wires, the resistivity ρ is $1.69 \times 10^{-8}\ \Omega\cdot\text{m}$, so the electric field associated with the current is

$$\begin{aligned} E = \rho j &= (1.69 \times 10^{-8}\ \Omega\cdot\text{m})(3 \times 10^5\text{ A/m}^2) \\ &\approx 5 \times 10^{-3}\text{ V/m}. \end{aligned}$$

This is a very small electric field.

When the battery is first connected to the circuit, initial transient currents are established. These currents distribute charges along the surfaces of the wires in just the precise way necessary to establish the electric field that maintains the steady current in the wires—and the entire process takes place in a time that is typically of the order of nanoseconds!

How much charge on the surface of the wire is needed to produce a field of 5×10^{-3} V/m in its interior? To find a rough order-of-magnitude estimate, we can use Eq. 26-6 for the electric field of a point charge, $E = q/4\pi\epsilon_0 R^2$, to find

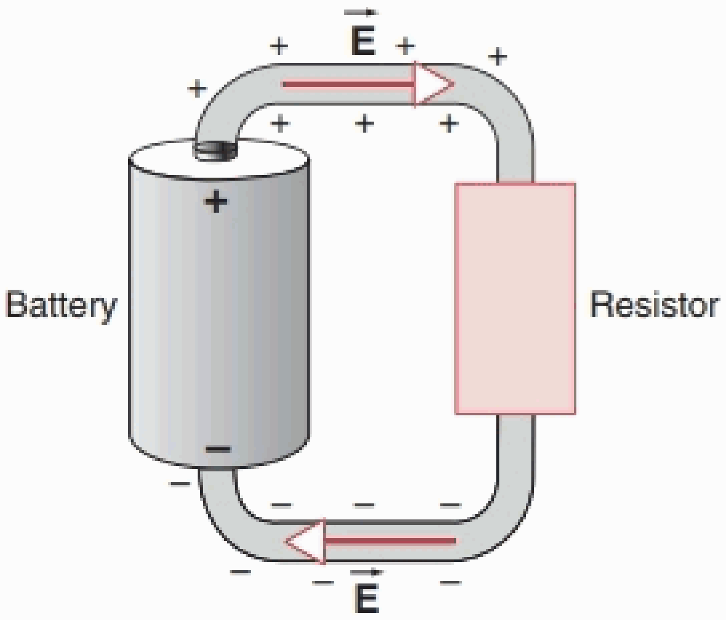


FIGURE 31-12. The electric fields in the connecting wires in a single-loop circuit. Charges on the surfaces of the wires are responsible for the fields in the wires.

---

*For a more detailed discussion of this topic, see *Electric and Magnetic Interactions* by R. Chabay and B. Sherwood (New York: Wiley, 1995), chapter 6. See also W. G. V. Rosser, *American Journal of Physics,* Vol. 31, 1963, p. 884 and Vol. 38, 1970, p. 265.

the charge q on the surface of the wire that would give a field E at its center:

$$q = 4\pi\epsilon_0 R^2 E = \frac{(10^{-3}\text{ m})^2(5 \times 10^{-3}\text{ V/m})}{9 \times 10^9\text{ N}\cdot\text{m}^2/\text{C}^2}$$

$$= 5.5 \times 10^{-19}\text{ C}$$

or about 3.5 electrons! We need only tiny amounts of charge on the surface to provide the electric field needed to sustain even a current as large as one ampere in a conductor.

In a similar fashion, we can ask how the current "knows" to change direction when it encounters a bend in the wire. Once again, the initial transients must result in just enough charge on the surface to guide the current. Figure 31-13 shows a schematic view of a right-angle bend, for which the surface charges must be distributed roughly as shown. The negative charge sets up a field near the bend that opposes the motion of the oncoming current, and the positive charge provides an initial "push" in the new direction. The preceding calculation again gives an order-of-magnitude estimate of the necessary charges—a few electrons on the surface are sufficient to change the direction of a one-ampere current!

Consideration of surface charge can also help us understand the effect of a resistor in a circuit. Let us consider a carbon resistor, and for simplicity we will give it the same diameter as the wires in our circuit (Fig. 31-14). Carbon is a poor conductor, but not quite a good insulator—its resistivity is about $3 \times 10^{-5}\ \Omega\cdot\text{m}$, about 2000 times larger than that of copper, but not nearly so large as that of typical insulators (see Table 29-1). Because the resistor and the wires have the same cross-sectional area, the current density is the same in both. Using our previous current density for the 1-A current in the 1-mm-radius wire, we can find the electric field in the resistor:

$$E = \rho j = (3 \times 10^{-5}\ \Omega\cdot\text{m})(3 \times 10^5\text{ A/m}^2) \approx 10\text{ V/m},$$

giving a field about 2000 times larger than the field in the copper wires. (Can you now see why the potential drop in the wires is negligible compared with the potential drop in the resistor?) This large electric field is necessary to force the electrons through the "constriction" in the circuit due to the resistor. As shown in Fig. 31-14, charges building up on the ends of the wire (similar to the charges on a capacitor) are responsible for producing this large electric field. For a resistor of thickness 1 mm, you should be able to show that about 1000 electrons on each end can produce the required field.

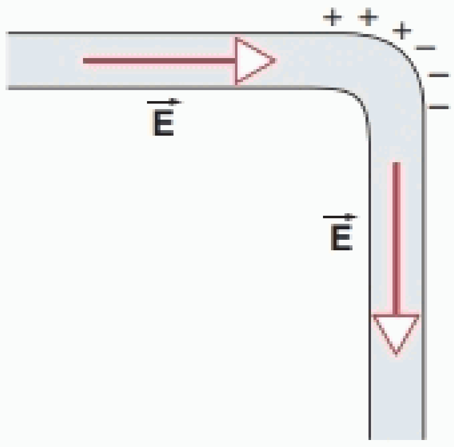


FIGURE 31-13. Detail of the surface charges near a right-angle bend.

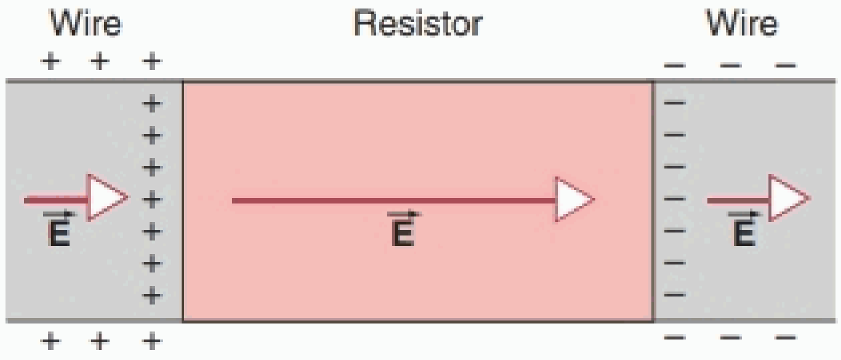


FIGURE 31-14. Details of surface charges near a resistor. The build-up of charge on the wire ends produces the large electric field in the resistor.

What we have described here is a wonderful, self-regulating system. The battery provides the initial "burst" of current to the circuit, and almost instantly the charge finds its way to the locations where it guides the steady current and prevents further build-up of charge on the surface of the wires. This equilibrium is maintained as long as the battery continues to pump charge around the circuit.

31-5 RESISTORS IN SERIES AND PARALLEL

As was the case with capacitors (see Section 30-4), resistors often occur in circuits in various combinations. In analyzing such circuits, it is helpful to replace the combination of resistors with a single *equivalent resistance* R_{eq}, whose value is chosen such that the operation of the circuit is unchanged. We consider two ways that resistors can be combined.

Resistors Connected in Parallel

Recall our definition of a parallel combination of circuit elements in Section 30-4: We can travel through the combination by crossing *only one* of the elements, the same potential difference ΔV appears across each element, and the flow of charge is shared among the elements.

Figure 31-15 shows two resistors connected in parallel. We seek the equivalent resistance between points a and b. Let us assume we connect a battery (or other source of emf) that maintains a potential difference ΔV between points a and b. The potential difference across each resistor

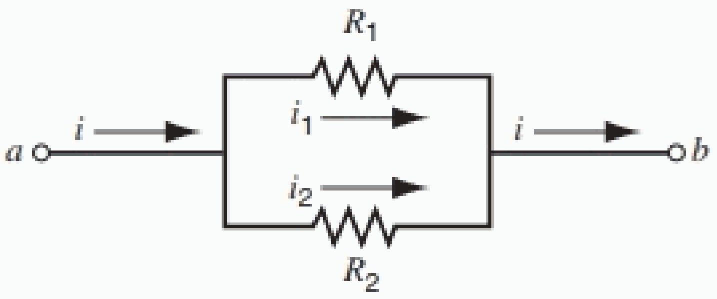


FIGURE 31-15. Two resistors in parallel.

is ΔV. The current through each of the resistors is, from Eq. 29-12,

$$i_1 = \Delta V/R_1 \quad \text{and} \quad i_2 = \Delta V/R_2. \tag{31-9}$$

According to the properties of a parallel circuit, the total current i must be shared among the branches, so

$$i = i_1 + i_2. \tag{31-10}$$

If we were to replace the parallel combination by a single equivalent resistance R_{eq}, the same total current i must flow (because the replacement must not change the operation of the circuit). The current is then

$$i = \Delta V/R_{\text{eq}}. \tag{31-11}$$

Substituting Eqs. 31-9 and 31-11 into Eq. 31-10, we obtain

$$\frac{\Delta V}{R_{\text{eq}}} = \frac{\Delta V}{R_1} + \frac{\Delta V}{R_2}$$

or

$$\frac{1}{R_{\text{eq}}} = \frac{1}{R_1} + \frac{1}{R_2}. \tag{31-12}$$

To find the equivalent resistance of a parallel combination of more than two resistors, we first find the equivalent resistance R_{12} of R_1 and R_2 using Eq. 31-12. We then find the equivalent resistance of R_{12} and the next parallel resistance, R_3, again using Eq. 31-12. Continuing in this way, we obtain a general expression for the equivalent resistance of a parallel combination of any number of resistors,

$$\frac{1}{R_{\text{eq}}} = \sum_n \frac{1}{R_n} \qquad \text{(parallel combination).} \tag{31-13}$$

That is, to find the equivalent resistance of a parallel combination, add the reciprocals of the individual resistances and take the reciprocal of the resulting sum. Note that R_{eq} is always *smaller than* the smallest resistance in the parallel combination—by adding more paths for the current, we get more current for the same potential difference.

In the special case of two resistors in parallel, Eq. 31-12 can be written

$$R_{\text{eq}} = \frac{R_1 R_2}{R_1 + R_2}, \tag{31-14}$$

or as the product of the two resistances divided by their sum.

Resistors Connected in Series

Figure 31-16 shows two resistors connected in series. Recall the properties of a series combination of circuit elements (see Section 30-4): to travel through the combination, we must travel through *all* the elements in succession, the potential difference across the combination is the sum of the potential differences across each element, and the same current is maintained in each element.

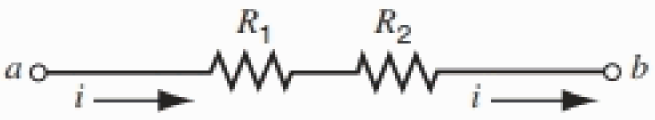


FIGURE 31-16. Two resistors in series.

Suppose a battery of potential difference ΔV is connected across points a and b in Fig. 31-16. A current i is set up in the combination and in each of the resistors. The potential differences across the resistors are

$$\Delta V_1 = iR_1 \quad \text{and} \quad \Delta V_2 = iR_2. \tag{31-15}$$

The sum of these potential differences must give the potential difference across points a and b maintained by the battery, or

$$\Delta V = \Delta V_1 + \Delta V_2. \tag{31-16}$$

If we replaced the combination by its equivalent resistance R_{eq}, the same current i would be established, and so

$$\Delta V = iR_{\text{eq}}. \tag{31-17}$$

Combining Eqs. 31-15, 31-16, and 31-17, we obtain

$$iR_{\text{eq}} = iR_1 + iR_2,$$

or

$$R_{\text{eq}} = R_1 + R_2. \tag{31-18}$$

Extending this result to a series combination of any number of resistors, we obtain

$$R_{\text{eq}} = \sum_n R_n \qquad \text{(series combination).} \tag{31-19}$$

That is, to find the equivalent resistance of a series combination, find the sum of the individual resistors. Note that the equivalent resistance of a series combination is always *larger* than the largest resistance in the series—adding more resistors in series means we get less current for the same potential difference.

Comparing these results with Eqs. 30-17 and 30-22 for the series and parallel combinations of capacitors, we see that resistors in parallel add like capacitors in series, and resistors in series add like capacitors in parallel. This has to do with the different way the two quantities are defined, resistance being potential/current and capacitance being charge/potential.

Occasionally, resistors may appear in combinations that are neither parallel nor series. In such a case, the equivalent resistance can sometimes be found by breaking the problem into smaller units that can be regarded as series or parallel connections, as the following sample problems demonstrate.

SAMPLE PROBLEM 31-5. (*a*) Find the equivalent resistance of the combination shown in Fig. 31-17*a*, using the values $R_1 = 4.6\ \Omega$, $R_2 = 3.5\ \Omega$, and $R_3 = 2.8\ \Omega$. (*b*) What is the value of the current through R_1 when a 12.0-V battery is connected across points a and b?

Solution (*a*) We first find the equivalent resistance R_{12} of the parallel combination of R_1 and R_2. Using Eq. 31-14 we obtain

$$R_{12} = \frac{R_1 R_2}{R_1 + R_2} = \frac{(4.6\ \Omega)(3.5\ \Omega)}{4.6\ \Omega + 3.5\ \Omega} = 2.0\ \Omega.$$

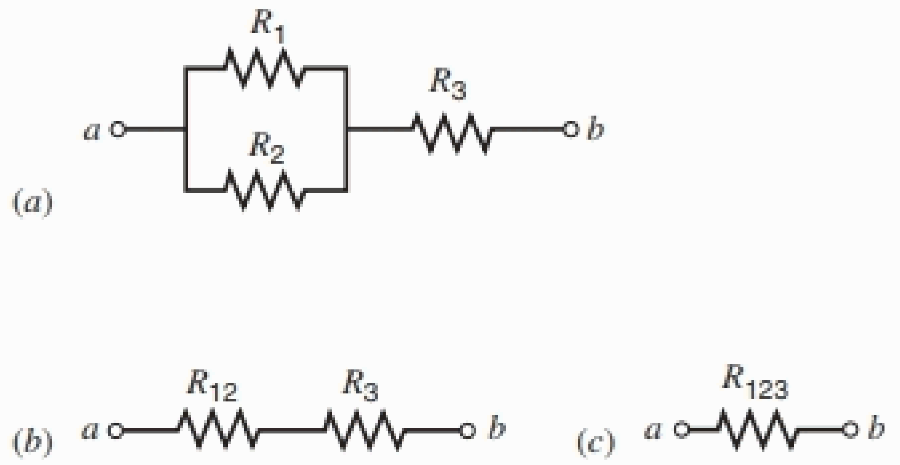


FIGURE 31-17. Sample Problem 31-5. (*a*) The parallel combination of R_1 and R_2 is in series with R_3. (*b*) The parallel combination of R_1 and R_2 has been replaced by its equivalent resistance, R_{12}. (*c*) The series combination of R_{12} and R_3 has been replaced by its equivalent resistance, R_{123}.

R_{12} and R_3 are in series, as shown in Fig. 31-17*b*. Using Eq. 31-18, we can find the equivalent resistance R_{123} of this series combination, which is the equivalent resistance of the entire original combination:

$$R_{123} = R_{12} + R_3 = 2.0\ \Omega + 2.8\ \Omega = 4.8\ \Omega.$$

(*b*) With a 12.0-V battery connected across points *a* and *b* in Fig. 31-17*c*, the resulting current is

$$i = \frac{\Delta V}{R_{123}} = \frac{12.0\ \text{V}}{4.8\ \Omega} = 2.5\ \text{A}.$$

With this current in the series combination in Fig. 31-17*b*, the potential difference across R_{12} is

$$\Delta V_{12} = iR_{12} = (2.5\ \text{A})(2.0\ \Omega) = 5.0\ \text{V}.$$

In a parallel combination, the same potential difference appears across each element (and across their combination). The potential difference across R_1 (and R_2) is therefore 5.0 V, and the current through R_1 is

$$i_1 = \frac{\Delta V_{12}}{R_1} = \frac{5.0\ \text{V}}{4.6\ \Omega} = 1.1\ \text{A}.$$

SAMPLE PROBLEM 31-6. Figure 31-18*a* shows a cube made of 12 resistors, each of resistance *R*. Find R_{12}, the equivalent resistance across a cube edge.

Solution Although this problem at first looks hopeless to divide into series and parallel subunits, the symmetry of the connections suggests a way to do so. The key is the realization that, from considerations of symmetry alone, points 3 and 6 must be at the same potential, and so must points 4 and 5.

If two points in a circuit have the same potential, the currents in the circuit do not change if you connect these points by a wire. There is no current in the wire because there is no potential difference between its ends. Points 3 and 6 may therefore be connected by a wire, and similarly points 4 and 5 may be connected.

This allows us to redraw the cube as in Fig. 31-18*b*. From this point, it is simply a matter of reducing the circuit between the input terminals to a single resistor, using the rules for resistors in series and in parallel. In Fig. 31-18*c*, we make a start by replacing five parallel combinations of two resistors by their equivalents, each of resistance $\frac{1}{2}R$.

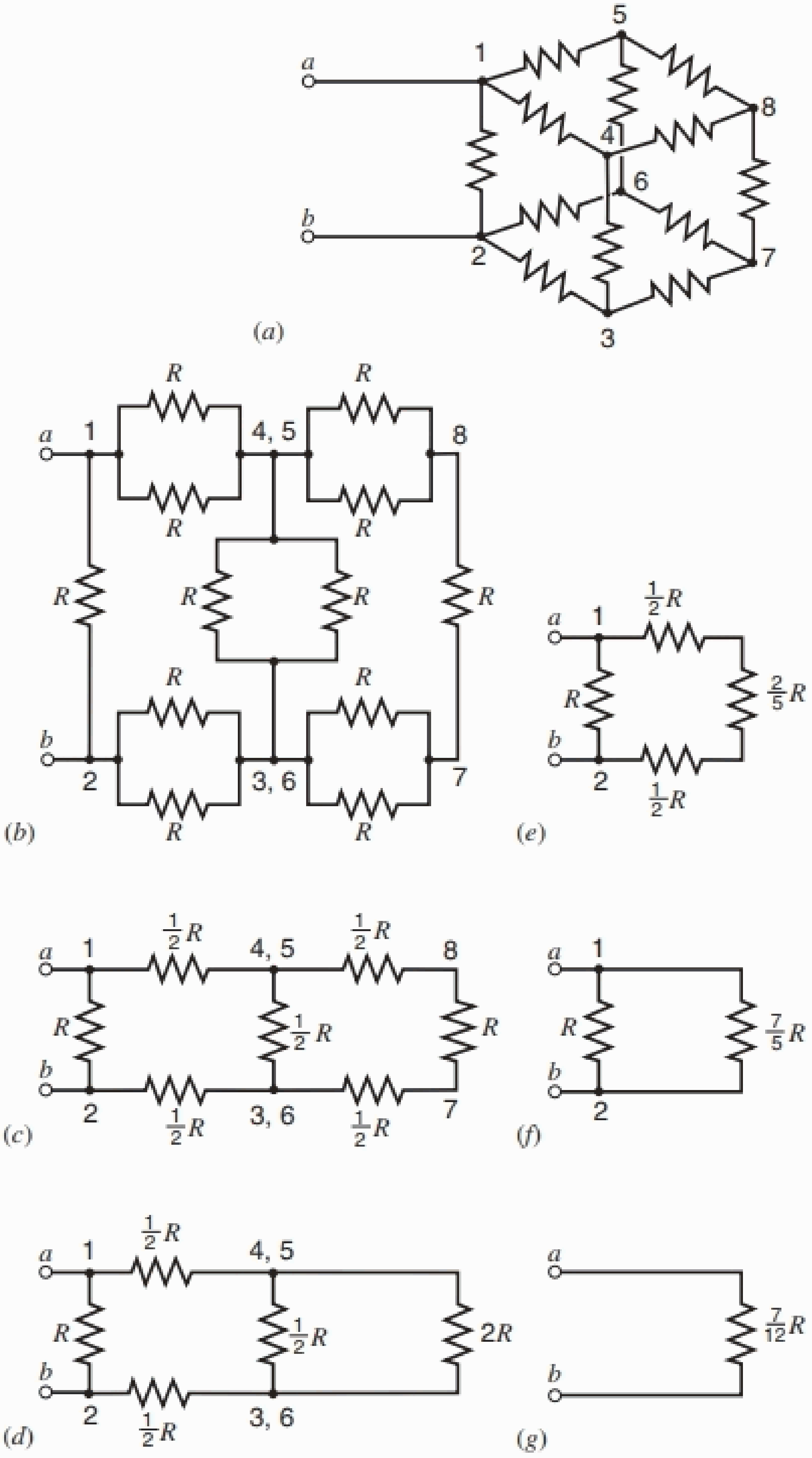


FIGURE 31-18. Sample Problem 31-6. (*a*) A cube formed of 12 identical resistors. (*b*)–(*g*) The step-by-step reduction of the cube to a single equivalent resistance.

In Fig. 31-18*d*, we have added the three resistors that are in series in the right-hand loop, obtaining a single equivalent resistance of 2*R*. In Fig. 31-18*e*, we have replaced the two resistors that now form the right-hand loop by a single equivalent resistor $\frac{2}{5}R$. In so doing, it is useful to recall that the equivalent resistance of two resistors in parallel is equal to their product divided by their sum (see Eq. 31-14).

In Fig. 31-18*f*, we have added the three series resistors of Fig. 31-18*e*, obtaining $\frac{7}{5}R$, and in Fig. 31-18*g* we have reduced this parallel combination to the single equivalent resistance that we seek—namely,

$$R_{12} = \tfrac{7}{12}R.$$

You can also use these methods to find R_{13}, the equivalent resistance of a cube across a face diagonal, and R_{17}, the equivalent resistance across a body diagonal (see Problem 7).

31-6 ENERGY TRANSFERS IN AN ELECTRIC CIRCUIT

Figure 31-19 shows a circuit consisting of a battery B connected to an electronic device, such as a resistor, a capacitor, a motor, or another battery. There is a current i in the wires and a potential difference ΔV_{ab} between the terminals of the device.

Let us first consider the operation of the battery, which we assume is an ideal (resistanceless) source of emf $\mathcal{E}$. As the battery moves a quantity of charge dq from its negative terminal to its positive terminal, it does work on the charge given by Eq. 31-1: $dW = \mathcal{E}\, dq$. The power delivered by the source of emf is determined by the rate at which work is being done; that is, $P_{\rm emf} = dW/dt = \mathcal{E}\, dq/dt$, or

$$P_{\rm emf} = \mathcal{E} i. \qquad (31\text{-}20)$$

This quantity gives the rate at which an ideal source of emf transfers energy to the rest of the circuit. As we discussed in Section 31-2, this energy might appear as the internal energy of a resistor, stored energy in the electric field of a capacitor, mechanical energy in a motor, or chemical energy in a battery that is being charged. If we consider the circuit to be an isolated system, then its total energy cannot change, and the decrease in energy of the source of emf must be balanced by an equivalent net increase in energy in the other parts of the circuit.

Suppose the circuit consists only of a source of emf and a resistor R. The potential difference between terminals a and b in Fig. 31-19 is $\Delta V_R = iR$. As a quantity of charge dq moves through the resistor from a to b, it experiences a potential energy change $dU = dq\ \Delta V_R$ (see Eq. 28-14). This energy must be transferred to the resistor, so the power transferred to the resistor (rate of energy transfer) is $P_R = dU/dt = (dq/dt)\ \Delta V_R = i\ \Delta V_R$ or

$$P_R = i^2 R. \qquad (31\text{-}21)$$

With $i = \Delta V_R/R$, we can also write this result as

$$P_R = \frac{(\Delta V_R)^2}{R}. \qquad (31\text{-}22)$$

This energy transfer to a resistor in a circuit is often known as *Joule heating.*

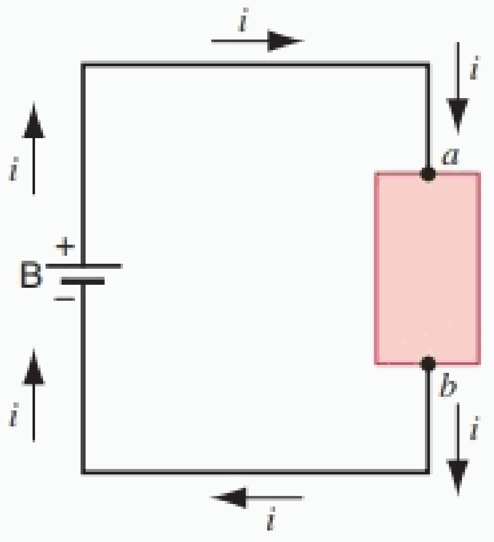


FIGURE 31-19. A battery B sets up a current i in a circuit containing an arbitrary electronic device.

As charge moves through the resistor from higher potential (terminal a) to lower potential (terminal b), it would tend to gain energy if it were not for the collisions with the atoms of the resistor. These collisions maintain the constant drift speed of the charge carriers, and the energy gained by the atoms in these collisions (which results in an increase in the amplitude of the vibration of the atoms about their equilibrium positions) can correspond to a temperature increase. This situation is analogous to the falling of a stone at its terminal velocity through a viscous medium such as air or water. As the stone falls in gravity, its decrease in potential energy is immediately transformed—not into an increase in its kinetic energy, but into an increase in the internal energy of the stone and the surrounding medium.

In a real battery with internal resistance r, the potential difference between the terminals is $\Delta V_{\rm batt} = \mathcal{E} - ir$, and the charge passing through the battery gains potential energy $dU = dq\ \Delta V_{\rm batt} = dq(\mathcal{E} - ir)$. The power delivered by this battery is $P_{\rm batt} = dU/dt$, or

$$P_{\rm batt} = \mathcal{E} i - i^2 r = P_{\rm emf} - P_r. \qquad (31\text{-}23)$$

The energy available to the rest of the circuit is decreased by the Joule heating in the internal resistance.

The unit of power that follows from Eqs. 31-20 to 31-22 is the volt · ampere, which you can show to be equivalent to the watt by using the definitions of the volt (joule/coulomb) and ampere (coulomb/second).

SAMPLE PROBLEM 31-7. You are given a length of heating wire made of a nickel–chromium–iron alloy called Nichrome; it has a resistance R of 72 Ω. It is to be connected across a 120-V line. Under which circumstances will the wire dissipate more heat: (*a*) its entire length is connected across the line, or (*b*) the wire is cut in half and the two halves are connected in parallel across the line?

Solution (*a*) The power P_R dissipated by the entire wire is, from Eq. 31-22,

$$P_R = \frac{(\Delta V)^2}{R} = \frac{(120\ \text{V})^2}{72\ \Omega} = 200\ \text{W}.$$

(*b*) The power for a wire of half length (and thus half resistance) is

$$P'_R = \frac{(\Delta V)^2}{\frac{1}{2}R} = \frac{(120\ \text{V})^2}{36\ \Omega} = 400\ \text{W}.$$

There are two halves, so the power obtained from both of them is 800 W, or four times that for the single wire. This would seem to suggest that you could buy a heating wire, cut it in half, and reconnect it to obtain four times the heat output. Why is this not such a good idea?

31-7 *RC* CIRCUITS

The preceding sections dealt with circuits containing only resistors, in which the currents did not vary with time. Here we introduce the capacitor as a circuit element, which leads us to the study of time-varying currents.

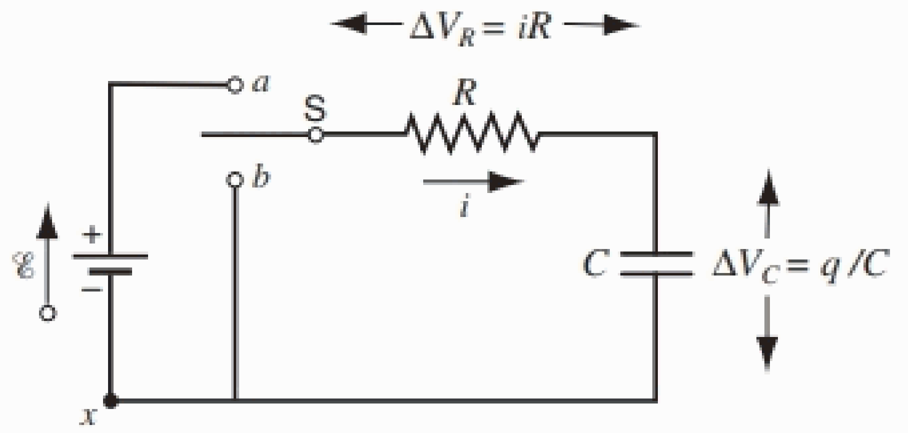


FIGURE 31-20. When switch S is connected to a, the capacitor C is charged by emf ℰ through the resistor R. After the capacitor is charged, the switch is moved to b, and the capacitor discharges through R. We can easily measure the potential difference ΔV_R $(= iR)$ across the resistor to determine the current i and also measure the potential difference ΔV_C $(= q/C)$ across the capacitor to determine the charge q.

Suppose we charge the capacitor in Fig. 31-20 by throwing switch S to position a. (Later we consider the connection to position b.) What current is set up in the resulting single-loop circuit? Let us apply conservation of energy principles.

In time dt a charge dq $(= i\,dt)$ moves through any cross section of the circuit and is deposited on the positive plate of the capacitor. The work ($= \mathscr{E}\,dq$; see Eq. 31-1) done by the source of emf must equal the internal energy $(= i^2R\,dt)$ produced in the resistor during time dt, plus the increase dU in the amount of energy U ($= q^2/2C$; see Eq. 30-24) that is stored in the capacitor. Conservation of energy gives

$$\mathscr{E}\,dq = i^2R\,dt + d\left(\frac{q^2}{2C}\right)$$

or

$$\mathscr{E}\,dq = i^2R\,dt + \frac{q}{C}\,dq.$$

Dividing by dt yields

$$\mathscr{E}\frac{dq}{dt} = i^2R + \frac{q}{C}\frac{dq}{dt}.$$

Since q is the charge on the upper plate, positive i means positive dq/dt. With $i = dq/dt$, this equation becomes

$$\mathscr{E} = iR + \frac{q}{C}. \qquad (31\text{-}24)$$

Equation 31-24 also follows from the loop rule, as it must, since the loop rule was derived from the conservation of energy principle. Starting from point x and going around the circuit clockwise, we experience an increase in potential in going through the source of emf and decreases in potential in going through the resistor and the capacitor, or

$$\mathscr{E} - iR - \frac{q}{C} = 0,$$

which is identical to Eq. 31-24.

To solve Eq. 31-24, we first substitute dq/dt for i, which gives

$$\mathscr{E} = R\frac{dq}{dt} + \frac{q}{C}. \qquad (31\text{-}25)$$

We can rewrite Eq. 31-25 as

$$\frac{dq}{q - \mathscr{E}C} = -\frac{dt}{RC}. \qquad (31\text{-}26)$$

Integrating this result in the case that $q = 0$ at $t = 0$, we obtain (after solving for q),

$$q = C\mathscr{E}(1 - e^{-t/RC}). \qquad (31\text{-}27)$$

We can check that this function $q(t)$ is really a solution of Eq. 31-25 by substituting it and its first derivative into that equation and seeing whether an identity results. Differentiating Eq. 31-27 with respect to time yields

$$i = \frac{dq}{dt} = \frac{\mathscr{E}}{R}e^{-t/RC}. \qquad (31\text{-}28)$$

Substituting q (Eq. 31-27) and dq/dt (Eq. 31-28) into Eq. 31-25 yields an identity, as you should verify. Equation 31-27 is therefore a solution of Eq. 31-25.

In the laboratory we can determine i and q conveniently by measuring quantities that are proportional to them—namely, the potential difference ΔV_R $(= iR)$ across the resistor and the potential difference ΔV_C $(= q/C)$ across the capacitor. Such measurements can be obtained rather easily by connecting voltmeters (or oscilloscope probes) across the resistor and the capacitor. Figure 31-21 shows the resulting plots of ΔV_R and ΔV_C. Note the following: (1) At $t = 0$, $\Delta V_R = \mathscr{E}$ (the full potential difference appears across R), and $\Delta V_C = 0$ (the capacitor is not charged). (2) As $t \to \infty$, $\Delta V_C \to \mathscr{E}$ (the capacitor becomes fully charged), and $\Delta V_R \to 0$ (the current stops flowing). (3) At all times, $\Delta V_R + \Delta V_C = \mathscr{E}$, as Eq. 31-25 requires.

The quantity RC in Eqs. 31-27 and 31-28 has the dimensions of time and is called the *capacitive time constant*

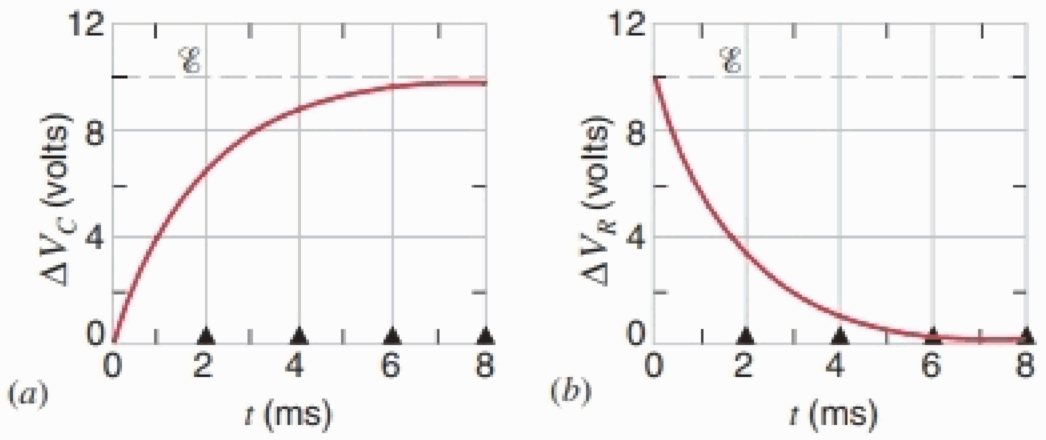


FIGURE 31-21. (a) As indicated by the potential difference ΔV_C, during the charging process the charge on the capacitor increases with time, and ΔV_C approaches the value of the emf ℰ. The time is measured after the switch is closed on a at $t = 0$. (b) The potential difference across the resistor decreases with time, approaching 0 at later times because the current falls to zero once the capacitor is fully charged. The curves have been drawn for $\mathscr{E} = 10$ V, $R = 2000\ \Omega$, and $C = 1\ \mu$F. The filled triangles represent successive RC time constants.

τ_C of the circuit:

$$\tau_C = RC. \tag{31-29}$$

It is the time at which the charge on the capacitor has increased to within a factor of $1 - e^{-1}$ ($\approx 63\%$) of its final value $C\mathscr{E}$. To show this, we put $t = \tau_C = RC$ in Eq. 31-27 to obtain

$$q = C\mathscr{E}(1 - e^{-1}) = 0.63C\mathscr{E}.$$

Figure 31-21*a* shows that if a resistance is included in a circuit with a charging capacitor, the increase of the charge of the capacitor toward its limiting value is *delayed* by a time characterized by the time constant RC. With no resistor present ($RC = 0$), the charge would rise immediately to its limiting value. Although we have shown that this time delay follows from an application of the loop rule to RC circuits, it is important to develop a physical understanding of the causes of the delay.

When switch S in Fig. 31-20 is closed on a, the charge on the capacitor is initially zero, so the potential difference across the capacitor is initially zero. At this time, Eq. 31-24 shows that $\mathscr{E} = iR$, and so $i = \mathscr{E}/R$ at $t = 0$. Because of this current, charge flows to the capacitor and the potential difference across the capacitor increases with time. Equation 31-24 now shows that, because the emf $\mathscr{E}$ is a constant, any increase in the potential difference across the capacitor must be balanced by a corresponding *decrease* in the potential difference across the resistor, with a similar decrease in the current. This decrease in the current means that the charge on the capacitor increases more slowly. This process continues until the current decreases to zero, at which time there is no potential drop across the resistor. The entire potential difference of the emf now appears across the capacitor, which is fully charged ($q = C\mathscr{E}$). Unless changes are made in the circuit, there is no further flow of charge. Review the derivations of Eqs. 31-27 and 31-28 and study Fig. 31-20 with the qualitative arguments of this paragraph in mind.

SAMPLE PROBLEM 31-8. A resistor R ($= 6.2\ \text{M}\Omega$) and a capacitor C ($= 2.4\ \mu\text{F}$) are connected in series, and a 12-V battery of negligible internal resistance is connected across their combination. (*a*) What is the capacitive time constant of this circuit? (*b*) At what time after the battery is connected does the potential difference across the capacitor equal 5.6 V?

Solution (*a*) From Eq. 31-29,

$$\tau_C = RC = (6.2 \times 10^6\ \Omega)(2.4 \times 10^{-6}\ \text{F}) = 15\ \text{s}.$$

(*b*) The potential difference across the capacitor is $\Delta V_C = q/C$, which according to Eq. 31-27 can be written

$$\Delta V_C = \frac{q}{C} = \mathscr{E}(1 - e^{-t/RC}).$$

Solving for t, we obtain (using $\tau_C = RC$)

$$t = -\tau_C \ln\left(1 - \frac{\Delta V_C}{\mathscr{E}}\right)$$

$$= -(15\ \text{s}) \ln\left(1 - \frac{5.6\ \text{V}}{12\ \text{V}}\right) = 9.4\ \text{s}.$$

As we found above, after a time τ_C ($= 15$ s), the potential difference across the capacitor is $0.63\mathscr{E} = 7.6$ V. It is reasonable that in a shorter time of 9.4 s, the potential difference across the capacitor reaches only the smaller value of 5.6 V.

Discharging a Capacitor

Assume now that the switch S in Fig. 31-20 has been in position a for a time that is much greater than RC. For all practical purposes, the capacitor is fully charged, and no charge is flowing. The switch S is then thrown to position b. How do the charge of the capacitor and the current vary with time?

With the switch S closed on b, the capacitor discharges through the resistor. There is no emf in the circuit and Eq. 31-24 for the circuit, with $\mathscr{E} = 0$, becomes simply

$$iR + \frac{q}{C} = 0. \tag{31-30}$$

Putting $i = dq/dt$ allows us to write the equation of the circuit (compare Eq. 31-25) as

$$R\frac{dq}{dt} + \frac{q}{C} = 0. \tag{31-31}$$

The solution is, as you may readily derive by integration (after writing $dq/q = -dt/RC$) and verify by substitution,

$$q = q_0 e^{-t/\tau_C}, \tag{31-32}$$

q_0 being the initial charge on the capacitor ($= \mathscr{E}C$, in our case). The capacitive time constant τ_C ($= RC$) appears in this expression for a discharging capacitor as well as in that for a charging capacitor (Eq. 31-27). We see that at a time such that $t = \tau_C = RC$, the capacitor charge is reduced to $q_0 e^{-1}$, which is about 37% of the initial charge q_0.

Differentiating Eq. 31-32, we find the current during discharge,

$$i = \frac{dq}{dt} = -\frac{q_0}{RC} e^{-t/\tau_C}. \tag{31-33}$$

The negative sign shows that the current is in the direction opposite to that shown in Fig. 31-20. This is as it should be, since the capacitor is discharging rather than charging. If the capacitor is originally fully charged, then $q_0 = C\mathscr{E}$, and we can write Eq. 31-33 as

$$i = -\frac{\mathscr{E}}{R} e^{-t/\tau_C}. \tag{31-34}$$

The initial current, found by setting $t = 0$ in Eq. 31-34, is $-\mathscr{E}/R$. This is reasonable because the initial potential difference across the resistor is $\mathscr{E}$.

The potential differences across R and C, which are, respectively, proportional to i and q, can again be measured as indicated in Fig. 31-20. Typical results are shown in Fig. 31-22. Note that, as suggested by Eq. 31-32, ΔV_C ($= q/C$) falls exponentially from its maximum value, which occurs

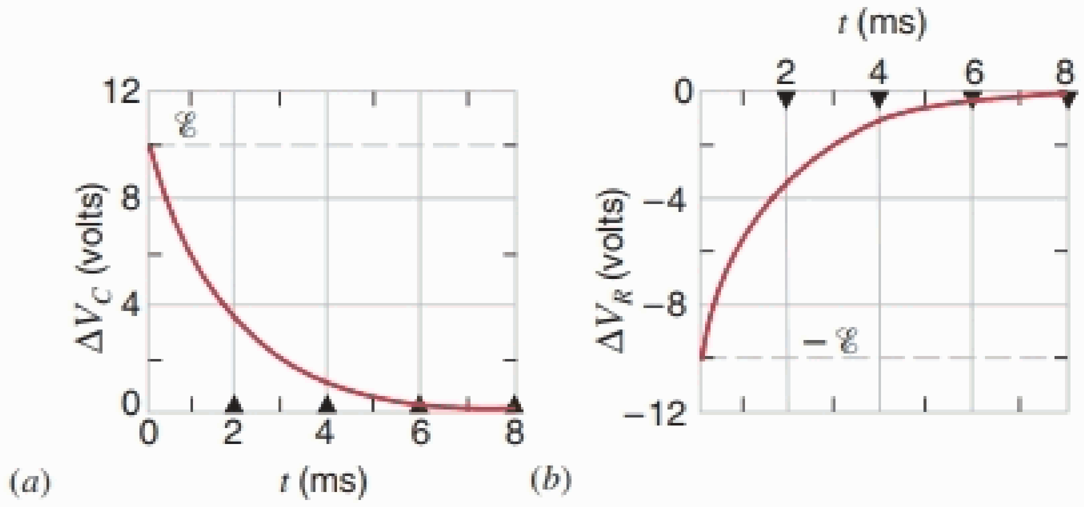


FIGURE 31-22. (*a*) After the capacitor has become fully charged, the switch in Fig. 31-20 is thrown from *a* to *b*, which we take to define a new $t = 0$. The potential difference across the capacitor decreases exponentially to zero as the capacitor discharges. (*b*) When the switch is initially moved to *b*, the potential difference across the resistor is negative compared with its value during the charging process shown in Fig. 31-21. As the capacitor discharges, the magnitude of the current falls exponentially to zero, and the potential drop across the resistor also approaches zero.

at $t = 0$, whereas ΔV_R $(= iR)$ is negative and rises exponentially to zero. Note also that $\Delta V_C + \Delta V_R = 0$, as required by Eq. 31-30.

SAMPLE PROBLEM 31-9. A capacitor C discharges through a resistor R. (*a*) After how many time constants does its charge fall to one-half its initial value? (*b*) After how many time constants does the stored energy drop to half its initial value?

Solution (*a*) The charge on the capacitor varies according to Eq. 31-32,

$$q = q_0 e^{-t/\tau_C},$$

in which q_0 is the initial charge. We seek the time t at which $q = \frac{1}{2}q_0$, or

$$\tfrac{1}{2}q_0 = q_0 e^{-t/\tau_C}.$$

Canceling q_0 and taking the natural logarithm of each side, we find

$$-\ln 2 = -\frac{t}{\tau_C}$$

or

$$t = (\ln 2)\tau_C = 0.69\tau_C.$$

The charge drops to half its initial value after 0.69 time constants. (*b*) The energy stored in the capacitor is

$$U = \frac{q^2}{2C} = \frac{q_0^2}{2C}e^{-2t/\tau_C} = U_0 e^{-2t/\tau_C},$$

in which U_0 is the initial stored energy. The time at which $U = \frac{1}{2}U_0$ is found from

$$\tfrac{1}{2}U_0 = U_0 e^{-2t/\tau_C}.$$

Canceling U_0 and taking the logarithm of each side, we obtain

$$-\ln 2 = -2t/\tau_C$$

or

$$t = \tau_C \frac{\ln 2}{2} = 0.35\tau_C.$$

The stored energy drops to half its initial value after 0.35 time constants have elapsed. This remains true no matter what the initial stored energy may be. The time $(0.69\tau_C)$ needed for the *charge* to fall to half its initial value is greater than the time $(0.35\tau_C)$ needed for the *energy* to fall to half its initial value. Why?

MULTIPLE CHOICE

31-1 Electric Current

1. The junction rule is a direct consequence of
   (A) Newton's second law.
   (B) conservation of momentum.
   (C) conservation of energy. (D) conservation of charge.

2. Figure 31-23 shows a network of wires carrying various currents. What is the current through A?
   (A) 1 A (B) 2 A (C) 3 A (D) 9 A
   (E) 11 A

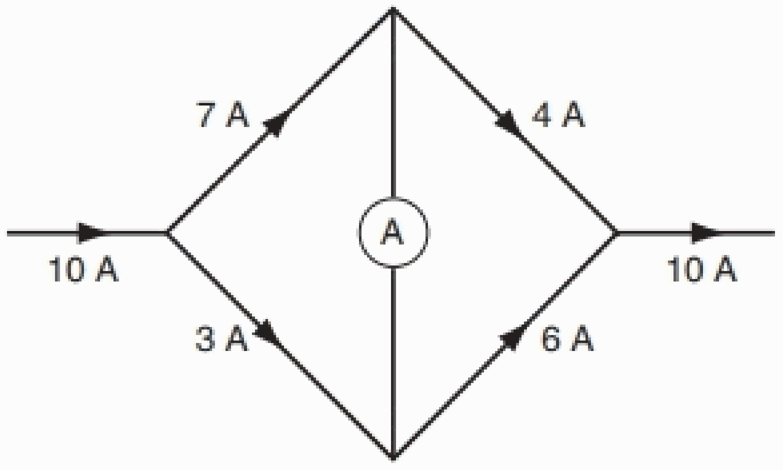


FIGURE 31-23. Multiple-choice question 2.

3. Figure 31-24 shows a network of wires carrying various currents. What is the current through A?
   (A) 1 A (B) 2 A (C) 6 A (D) 8 A
   (E) There is not enough information given.

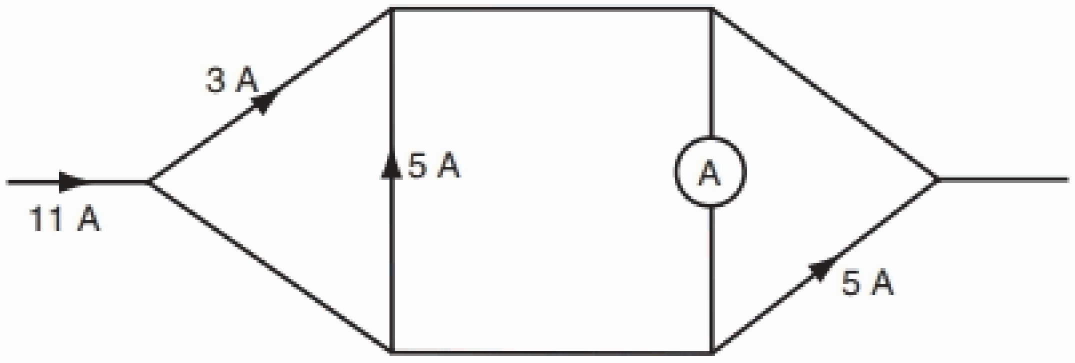


FIGURE 31-24. Multiple-choice question 3.

31-2 Electromotive Force

4. What are the units for ℰ, the electromotive force?
   (A) Emfs (B) Joules (C) Volts (D) Newtons

5. The function of the source of emf in a circuit is to
   (A) supply electrons to the circuit.
   (B) raise electrons to a higher potential.

(C) push electrons to a lower potential.
(D) accelerate electrons to greater speeds.

31-3 Analysis of Circuits

6. The loop rule is a direct consequence of
(A) Newton's second law.
(B) conservation of momentum.
(C) conservation of energy.
(D) conservation of charge.

7. A fixed resistor R is in series with a variable resistor and an ideal battery. Originally the resistances are the same.

(*a*) As the resistance of the variable resistor is decreased, the current through the variable resistor
(A) increases. (B) decreases.
(C) remains the same.
(D) cannot be determined without more information.

(*b*) As the resistance of the variable resistor is decreased, the potential difference across the variable resistor
(A) increases. (B) decreases.
(C) remains the same.
(D) cannot be determined without more information.

31-4 Electric Fields in Circuits

31-5 Resistors in Series and Parallel

8. Two resistors R_1 and R_2 are connected in series; assume that $R_1 < R_2$. The equivalent resistance of this arrangement is R, where
(A) $R < R_1/2$. (B) $R_1/2 < R < R_1$.
(C) $R_1 < R < R_2$. (D) $R_2 < R < 2R_2$.
(E) $2R_2 < R$.

9. Two resistors R_1 and R_2 are connected in parallel; assume that $R_1 < R_2$. The equivalent resistance of this argument is R, where
(A) $R < R_1/2$. (B) $R_1/2 < R < R_1$.
(C) $R_1 < R < R_2$. (D) $R_2 < R < 2R_2$.
(E) $2R_2 < R$.

10. What is the minimum number of resistors required to construct a network that *cannot* be analyzed to find the equivalent resistance by treating resistances in parallel or in series according to Eqs. 31-13 and 31-19?
(A) 3 (B) 4 (C) 5 (D) 6

31-6 Energy Transfers in an Electric Circuit

11. A standard lightbulb in the United States is the 60-watt bulb, which is designed to operate in a 120-volt circuit. During a brown-out it was noticed that the power output from the bulb decreased to 30 W. To what percentage of the original value did the voltage decrease?
(A) 75% (B) 70% (C) 50% (D) 33%

12. A fixed resistor R is in series with a variable resistor and a real battery (internal resistance is not negligible). Originally the fixed and variable resistors have the same resistance.

(*a*) As the resistance of the variable resistor is decreased, the rate at which energy is transferred to the fixed resistor
(A) increases. (B) decreases.
(C) remains the same.
(D) cannot be determined without more information.

(*b*) As the resistance of the variable resistor is decreased, the rate at which energy is transferred to the variable resistor
(A) increases. (B) decreases.
(C) remains the same.
(D) cannot be determined without more information.

13. A fixed resistor is in parallel with a variable resistor; both are connected to a real battery (internal resistance is not negligible). Originally the fixed and variable resistors have the same resistance.

(*a*) As the resistance of the variable resistor is decreased, the current through the fixed resistor
(A) increases. (B) decreases.
(C) remains the same.
(D) cannot be determined without more information.

(*b*) As the resistance of the variable resistor is decreased, the rate at which energy is transferred to the fixed resistor
(A) increases. (B) decreases.
(C) remains the same.
(D) cannot be determined without more information.

(*c*) As the resistance of the variable resistor is decreased, the rate at which energy is transferred to the variable resistor
(A) increases. (B) decreases.
(C) remains the same.
(D) cannot be determined without more information.

31-7 *RC* Circuits

14. A resistor, capacitor, switch, and ideal battery are in series. Originally the capacitor is uncharged. The switch is then closed, allowing current to flow.

(*a*) While the current is flowing, the potential difference across the resistor is
(A) increasing. (B) decreasing. (C) fixed.

(*b*) While the current is flowing, the potential difference across the capacitor is
(A) increasing. (B) decreasing. (C) fixed.

15. A capacitor is charged by connecting it in series with a resistor and an ideal battery. The battery supplies energy at a rate $P(t)$, the internal energy of the resistor increases at a rate $P_R(t)$, and the capacitor stores energy at a rate $P_C(t)$. What can be concluded about the relationship between $P_R(t)$ and $P_C(t)$?
(A) $P_R(t) > P_C(t)$ for all times t while charging.
(B) $P_R(t) = P_C(t)$ for all times t while charging.
(C) $P_R(t) < P_C(t)$ for all times t while charging.
(D) $P_R(t) > P_C(t)$ only at the beginning of the charging.
(E) $P_R(t) < P_C(t)$ only at the beginning of the charging.

QUESTIONS

1. Does the direction of the emf provided by a battery depend on the direction of current flow through the battery?

2. In Fig. 31-4, discuss what changes would occur if we increased the mass m by such an amount that the "motor" reversed direction and became a "generator"—that is, a source of emf.

3. Discuss in detail the statement that the energy method and the loop rule method for solving circuits are perfectly equivalent.

4. Devise a method for measuring the emf and the internal resistance of a battery.

5. What is the origin of the internal resistance of a battery? Does this depend on the age or size of the battery?

6. The current passing through a battery of emf $\mathcal{E}$ and internal resistance r is decreased by some external means. Does the potential difference between the terminals of the battery necessarily decrease or increase? Explain.

7. How could you calculate ΔV_{ab} in Fig. 31-8*a* by following a path from *a* to *b* that does not lie in the conducting circuit?

8. A 25-W, 120-V bulb glows at normal brightness when connected across a bank of batteries. A 500-W, 120-V bulb glows only dimly when connected across the same bank. How could this happen?

9. Under what circumstances can the terminal potential difference of a battery exceed its emf?

10. Automobiles generally use a 12-V electrical system. Years ago a 6-V system was used. Why the change? Why not 24 V?

11. The loop rule is based on the conservation of energy principle and the junction rule is based on the conservation of charge principle. Explain just how these rules are based on these principles.

12. Under what circumstances would you want to connect batteries in parallel? In series?

13. Compare and contrast the formulas for the equivalent values of series and parallel combinations of (*a*) capacitors and (*b*) resistors.

14. Under what circumstances would you want to connect resistors in parallel? In series?

15. What is the difference between an emf and a potential difference?

16. Referring to Fig. 31-10, use a qualitative argument to convince yourself that i_2 is drawn in the wrong direction.

17. Do the junction and loop rules apply to a circuit containing a capacitor?

18. Show that the product *RC* in Eq. 31-29 has the dimensions of time—that is, that 1 second = 1 ohm × 1 farad.

19. A capacitor, resistor, and battery are connected in series. The charge that the capacitor stores is unaffected by the resistance of the resistor. What purpose, then, is served by the resistor?

20. Explain why, in Sample Problem 31-9, the stored energy falls to half its initial value more rapidly than does the charge.

21. The light flash in a camera is produced by the discharge of a capacitor across the lamp. Why do we not just connect the photoflash lamp directly across the power supply used to charge the capacitor?

22. Does the time required for the charge on a capacitor in an *RC* circuit to build up to a given fraction of its final value depend on the value of the applied emf? Does the time required for the charge to change by a given amount depend on the applied emf?

23. A capacitor is connected across the terminals of a battery. Does the charge that eventually appears on the capacitor plates depend on the value of the internal resistance of the battery?

24. Devise a method whereby an *RC* circuit can be used to measure very high resistances.

25. In Fig. 31-20, suppose that switch S is closed on *a*. Explain why, in view of the fact that the negative terminal of the battery is not connected to resistance *R*, the current in *R* should be $\mathcal{E}/R$, as Eq. 31-28 predicts.

26. In Fig. 31-20, suppose that switch S is closed on *a*. Why does the charge on capacitor *C* not rise instantaneously to $q = C\mathcal{E}$? After all, the positive battery terminal is connected to one plate of the capacitor and the negative terminal to the other.

27. What special characteristics must heating wire have?

28. Equation 31-21 seems to suggest that the rate of increase of internal energy in a resistor is reduced if the resistance is made less; Eq. 31-22 seems to suggest just the opposite. How do you reconcile this apparent paradox?

29. Why do electric power companies reduce voltage during times of heavy demand? What is being saved?

30. Is the filament resistance lower or higher in a 500-W lightbulb than in a 100-W bulb? Both bulbs are designed to operate at 120 V.

31. Five wires of the same length and diameter are connected in turn between two points maintained at constant potential difference. Will internal energy be developed at a faster rate in the wire of (*a*) the smallest or (*b*) the largest resistance?

32. Why is it better to send 10 MW of electric power long distances at 10 kV rather than at 220 V?

EXERCISES

31-1 Electric Current

31-2 Electromotive Force

1. A 5.12-A current is set up in an external circuit by a 6.00-V storage battery for 5.75 min. By how much is the chemical energy of the battery reduced?

2. (*a*) How much work does a 12.0-V source of emf do on an electron as it passes through from the positive to the negative terminal? (*b*) If 3.40×10^{18} electrons pass through each second, what is the power output of the source?

3. A certain 12-V car battery can "pump" a total charge of 125 A · h before it runs down. Assuming that the potential difference across the terminals stays constant, for how long can it deliver energy at the rate of 110 W?

4. A standard flashlight battery can deliver about 2.0 W · h of energy before it runs down. (*a*) If a battery costs 80 cents, what is the cost of operating a 100-W lamp for 8.0 h using batteries? (*b*) What is the cost if power provided by an electric utility company, at 12 cents per kW · h, is used?

31-3 Analysis of Circuits

5. In Fig. 31-25, the potential at point P is 100 V. What is the potential at point Q?

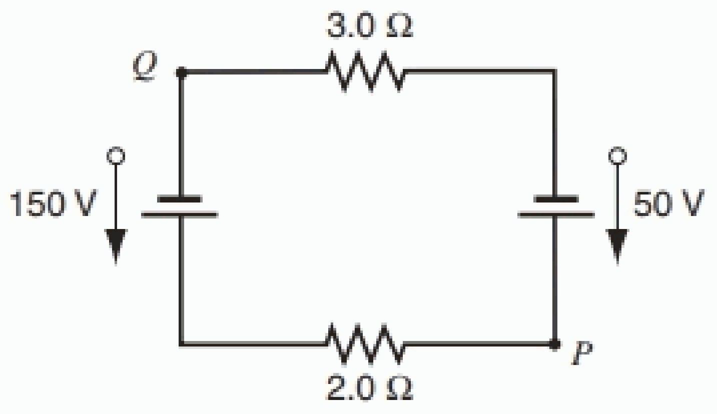


FIGURE 31-25. Exercise 5.

6. A gasoline gauge for an automobile is shown schematically in Fig. 31-26. The indicator (on the dashboard) has a resistance of 10 Ω. The tank unit is simply a float connected to a resistor that has a resistance of 140 Ω when the tank is empty, 20 Ω when it is full, and varies linearly with the volume of gasoline. Find the current in the circuit when the tank is (*a*) empty, (*b*) half full, and (*c*) full.

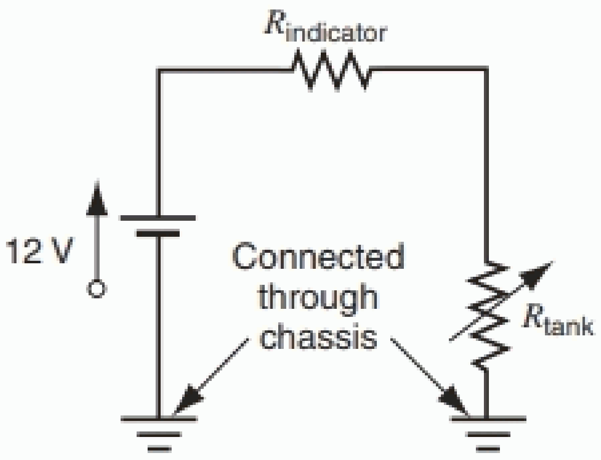


FIGURE 31-26. Exercise 6.

7. (*a*) In Fig. 31-27, what value must R have if the current in the circuit is to be 50 mA? Take $\mathscr{E}_1 = 2.0$ V, $\mathscr{E}_2 = 3.0$ V, and $r_1 = r_2 = 3.0\ \Omega$. (*b*) What is the rate at which internal energy appears in R?

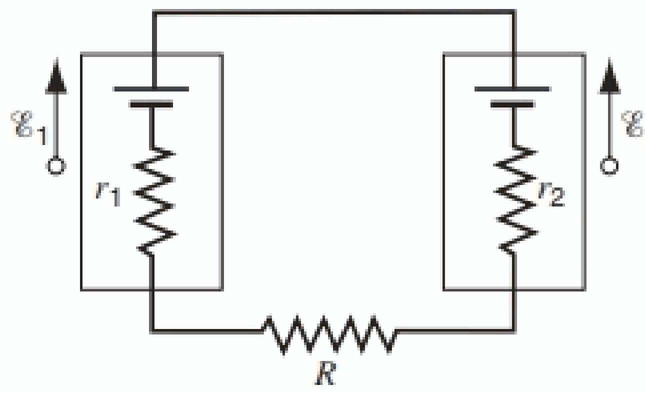


FIGURE 31-27. Exercise 7.

8. The current in a single-loop circuit is 5.0 A. When an additional resistance of 2.0 Ω is inserted in series, the current drops to 4.0 A. What was the resistance in the original circuit?

9. The section of circuit AB (see Fig. 31-28) absorbs 53.0 W of power when a current $i = 1.20$ A passes through it in the indicated direction. (*a*) Find the potential difference between A and B. (*b*) If the element C does not have an internal resistance, what is its emf? (*c*) Which terminal, left or right, is positive?

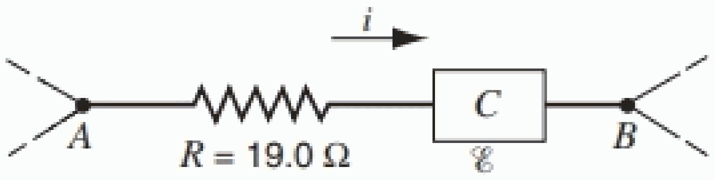


FIGURE 31-28. Exercise 9.

10. Internal energy is to be generated in a 108-mΩ resistor at the rate of 9.88 W by connecting it to a battery whose emf is 1.50 V. (*a*) What is the internal resistance of the battery? (*b*) What potential difference exists across the resistor?

11. What current, in terms of $\mathscr{E}$ and R, does the ammeter A in Fig. 31-29 read? Assume that A has zero resistance.

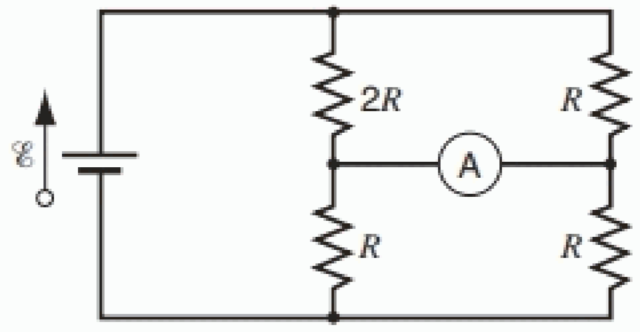


FIGURE 31-29. Exercise 11.

12. You are given two batteries of emf values $\mathscr{E}_1$ and $\mathscr{E}_2$ and internal resistances r_1 and r_2. They may be connected either in (*a*) parallel or (*b*) series and are used to establish a current in a resistor R, as shown in Fig. 31-30. Derive expressions for the current in R for both methods of connection.

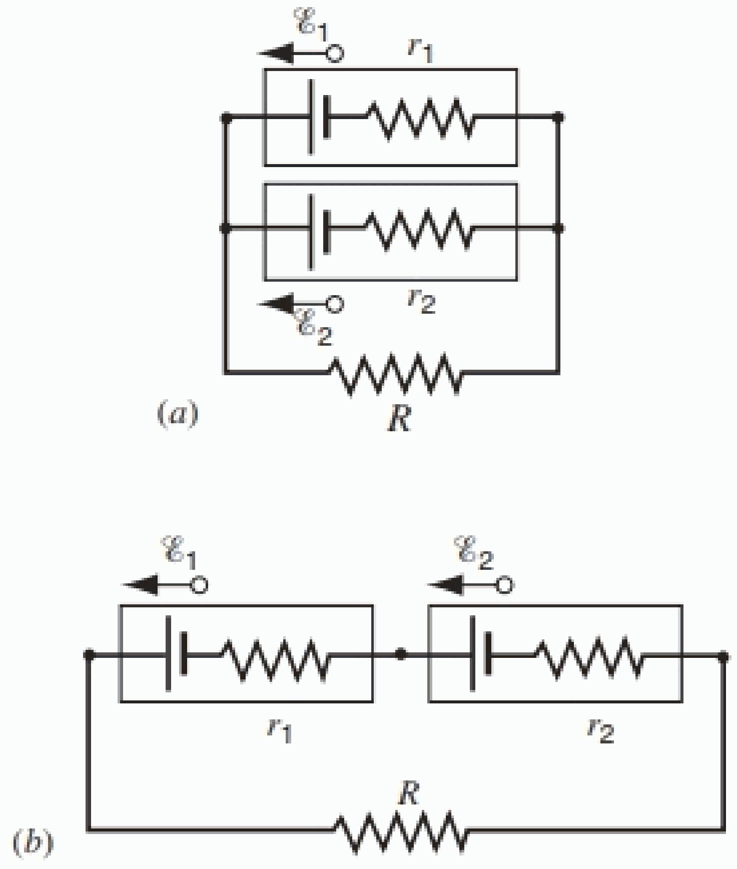


FIGURE 31-30. Exercise 12.

13. (*a*) Calculate the current through each source of emf in Fig. 31-31. (*b*) Calculate $V_b - V_a$. Assume that $R_1 = 1.20\ \Omega$, $R_2 = 2.30\ \Omega$, $\mathcal{E}_1 = 2.00$ V, $\mathcal{E}_2 = 3.80$ V, and $\mathcal{E}_3 = 5.00$ V.

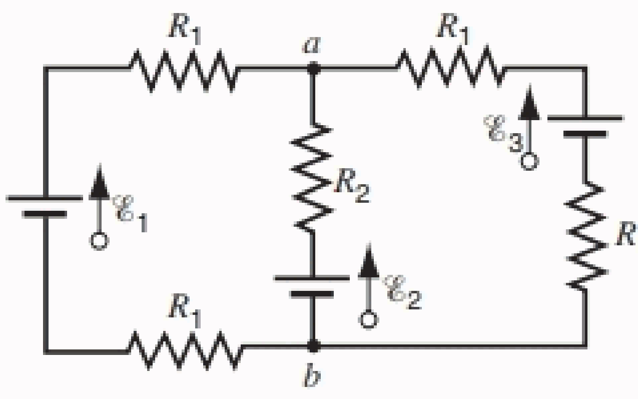


FIGURE 31-31. Exercise 13.

14. (*a*) In the circuit of Fig. 31-8, show that the power delivered to R as internal energy is a maximum when R is equal to the internal resistance r of the battery. (*b*) Show that this maximum power is $P = \mathcal{E}^2/4r$.

15. A battery of emf $\mathcal{E} = 2.0$ V and internal resistance $r = 0.50\ \Omega$ is driving a motor. The motor is lifting a 2.0-N object at constant speed $v = 0.50$ m/s. Assuming no power losses, find (*a*) the current i in the circuit and (*b*) the potential difference ΔV across the terminals of the motor. (*c*) Discuss the fact that there are two solutions to this problem.

31-4 Electric Fields in Circuits

31-5 Resistors in Series and Parallel

16. Four 18-Ω resistors are connected in parallel across a 27-V battery. What is the current through the battery?

17. By using only two resistors—singly, in series, or in parallel—you are able to obtain resistances of 3.0 Ω, 4.0 Ω, 12 Ω, and 16 Ω. What are the separate resistances of the resistors?

18. In Fig. 31-32, find the equivalent resistance between points (*a*) A and B, (*b*) A and C, and (*c*) B and C.

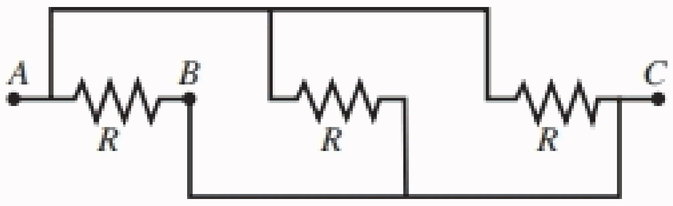


FIGURE 31-32. Exercise 18.

19. A circuit containing five resistors connected to a 12-V battery is shown in Fig. 31-33. Find the potential difference across the 5.0-Ω resistor.

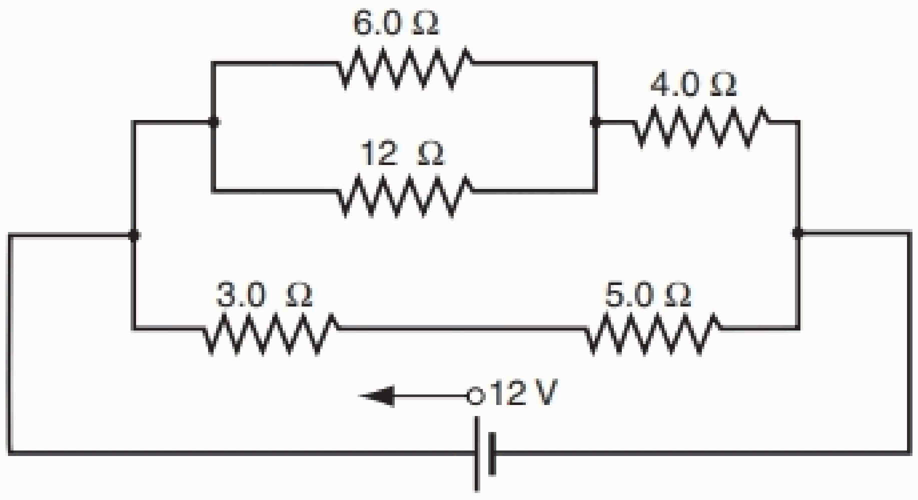


FIGURE 31-33. Exercise 19.

20. A 120-V power line is protected by a 15-A fuse. What is the maximum number of 500-W lamps that can be simultaneously operated in parallel on this line?

21. Two resistors R_1 and R_2 may be connected either in series or parallel across a (resistanceless) battery with emf $\mathcal{E}$. We desire the internal energy transfer rate for the parallel combination to be five times that for the series combination. If $R_1 = 100\ \Omega$, what is R_2?

22. You are given a number of 10-Ω resistors, each capable of dissipating only 1.0 W. What is the minimum number of such resistors that you need to combine in series or parallel combinations to make a 10-Ω resistor capable of dissipating at least 5.0 W?

23. (*a*) In Fig. 31-34, find the equivalent resistance of the network shown. (*b*) Calculate the current in each resistor. Put $R_1 = 112\ \Omega$, $R_2 = 42.0\ \Omega$, $R_3 = 61.6\ \Omega$, $R_4 = 75.0\ \Omega$, and $\mathcal{E} = 6.22$ V.

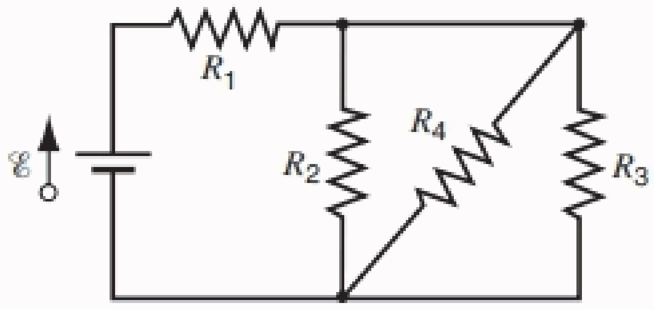


FIGURE 31-34. Exercise 23.

24. In the circuit of Fig. 31-35, $\mathcal{E}$, R_1, and R_2 have constant values but R can be varied. Find an expression for R that results in the maximum power dissipation in that resistor.

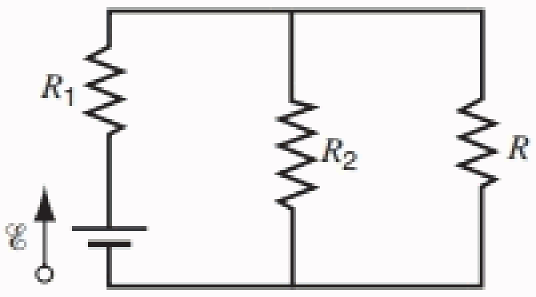


FIGURE 31-35. Exercise 24.

25. In Fig. 31-36, find the equivalent resistance between points (*a*) F and H and (*b*) F and G.

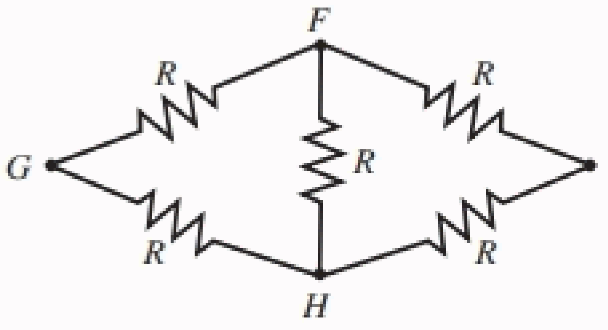


FIGURE 31-36. Exercise 25.

26. A voltage divider consists of two resistors in series. The applied potential difference across the resistors is 12 V, and the potential difference across the second resistor is 2.4 V. Find the resistances, assuming that the current through the two resistors is 1 mA.

27. Design a voltage divider that will input 1.50 V and output approximately 0.95 ± 0.01 V using only standard resistor values.

28. A portion of an infinite array of identical 1-$\mu\Omega$ resistors is shown in Fig. 31-37. A battery is connected across two distant junctions. Show that the potential at any junction is the average of the potential at the four nearest junctions. You will use this result to solve Computer Problem 1.

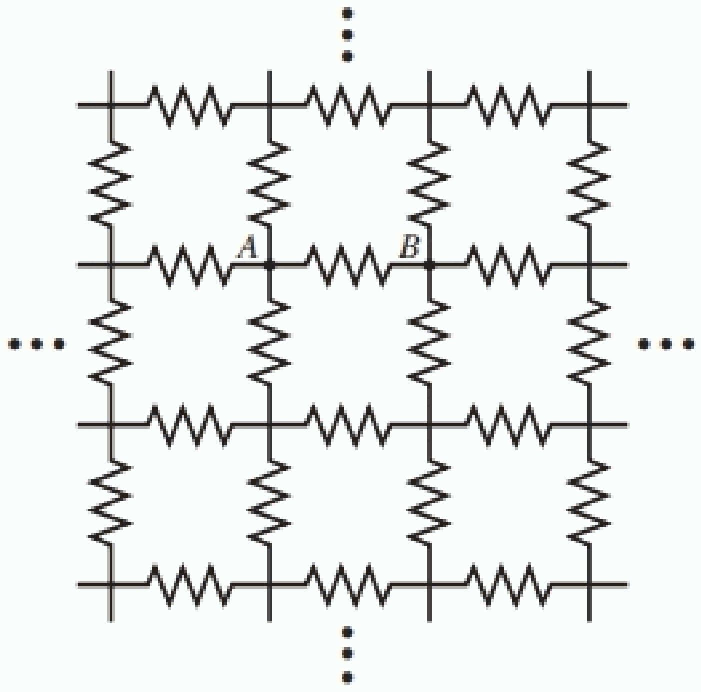


FIGURE 31-37. Exercise 28 and Computer Problem 1.

31-6 Energy Transfers in an Electric Circuit

29. A student's 9.0-V, 7.5-W portable radio was left on from 9:00 P.M. until 3:00 A.M. How much charge passed through the wires?

30. The headlights of a moving car draw 9.7 A from the 12-V alternator, which is driven by the engine. Assume that the alternator is 82% efficient, and calculate the horsepower the engine must supply to run the lights.

31. A space heater, operating from a 120-V line, has a hot resistance of 14.0 Ω. (*a*) At what rate is electrical energy transferred into internal energy? (*b*) At 5.22¢/kW·h, what does it cost to operate the device for 6 h 25 min?

32. Figure 31-38 shows a battery connected across a uniform resistor R_0. A sliding contact can move across the resistor from $x = 0$ at the left to $x = 10$ cm at the right. Find an expression for the power dissipated in the resistor R as a function of x. Plot the function for $\mathcal{E} = 50$ V, $R = 2000\ \Omega$, and $R_0 = 100\ \Omega$.

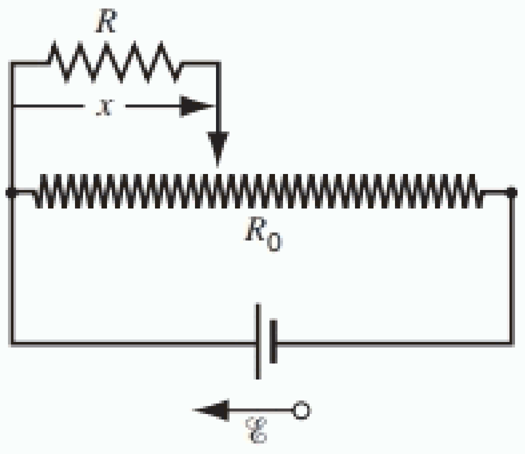


FIGURE 31-38. Exercise 32.

33. Two lightbulbs, one of resistance R_1 and the other of resistance R_2 ($< R_1$), are connected (*a*) in parallel and (*b*) in series. Which bulb is brighter in each case?

34. The National Board of Fire Underwriters has fixed safe current-carrying capacities for various sizes and types of wire. For #10 rubber-coated copper wire (diameter = 0.10 in.), the maximum safe current is 25 A. At this current, find (*a*) the current density, (*b*) the electric field, (*c*) the potential difference for 1000 ft of wire, and (*d*) the rate at which internal energy is developed for 1000 ft of wire.

35. A 100-W lightbulb is plugged into a standard 120-V outlet. (*a*) How much does it cost per month (31 days) to leave the light turned on? Assume electric energy costs 6¢/kW·h. (*b*) What is the resistance of the bulb? (*c*) What is the current in the bulb? (*d*) Is the resistance different when the bulb is turned off?

36. A Nichrome heater dissipates 500 W when the applied potential difference is 110 V and the wire temperature is 800°C. How much power would it dissipate if the wire temperature were held at 200°C by immersion in a bath of cooling oil? The applied potential difference remains the same; α for Nichrome at 800°C is 4.0×10^{-4}/C°.

37. An electron linear accelerator produces a pulsed beam of electrons. The pulse current is 485 mA and the pulse duration is 95.0 ns. (*a*) How many electrons are accelerated per pulse? (*b*) Find the average current for a machine operating at 520 pulses/s. (*c*) If the electrons are accelerated to an energy of 47.7 MeV, what are the values of average and peak power outputs of the accelerator?

38. A cylindrical resistor of radius 5.12 mm and length 1.96 cm is made of material that has a resistivity of 3.50×10^{-5} Ω·m. What are (*a*) the current density and (*b*) the potential difference when the power dissipation is 1.55 W?

39. A heating element is made by maintaining a potential difference of 75 V along the length of a Nichrome wire with a 2.6-mm^2 cross section and a resistivity of 5.0×10^{-7} Ω·m. (*a*) If the element dissipates 4.8 kW, what is its length? (*b*) If a potential difference of 110 V is used to obtain the same power output, what should the length be?

40. A 420-W immersion heater is placed in a pot containing 2.10 liters of water at 18.5°C. (*a*) How long will it take to bring the water to boiling temperature, assuming that 77.0% of the available energy is absorbed by the water? (*b*) How much longer will it take to boil half the water away?

31-7 *RC* Circuits

41. (*a*) Carry out the missing steps to obtain Eq. 31-27 from Eq. 31-26. (*b*) In a similar manner, obtain Eq. 31-32 from Eq. 31-31. Note that $q = q_0$ (capacitor charged) at $t = 0$.

42. In an *RC* series circuit $\mathcal{E} = 11.0$ V, $R = 1.42$ MΩ, and $C = 1.80\ \mu$F. (*a*) Calculate the time constant. (*b*) Find the maximum charge that will appear on the capacitor during charging. (*c*) How long does it take for the charge to build up to 15.5 μC?

43. How many time constants must elapse before a capacitor in an *RC* circuit is charged to 99% of its maximum charge?

44. A 15.2-kΩ resistor and a capacitor are connected in series and a 13.0-V potential is suddenly applied. The potential across the capacitor rises to 5.00 V in 1.28 μs. (*a*) Calculate the time constant. (*b*) Find the capacitance of the capacitor.

45. An *RC* circuit is discharged by closing a switch at time $t = 0$. The initial potential difference across the capacitor is 100 V. (*a*) If the potential difference has decreased to 1.06 V after 10.0 s, calculate the time constant of the circuit. (*b*) What will be the potential difference at $t = 17$ s?

46. A controller on an electronic arcade game consists of a variable resistor connected across the plates of a 220-nF capacitor. The capacitor is charged to 5.00 V, then discharged through the resistor. The time for the potential difference across the plates to decrease to 800 mV is measured by an internal clock. If the range of discharge times that can be handled is from 10.0 μs to 6.00 ms, what should be the range of the resistance of the resistor?

47. Figure 31-39 shows the circuit of a flashing lamp, like those attached to barrels at highway construction sites. The fluorescent lamp *L* is connected in parallel across the capacitor *C* of an *RC* circuit. Current passes through the lamp only when the potential across it reaches the breakdown voltage V_L; in this event, the capacitor discharges through the lamp and it flashes for a very short time. Suppose that two flashes per second are needed. Using a lamp with breakdown voltage $V_L = 72$ V, a 95-V battery, and a 0.15-μF capacitor, what should be the resistance *R* of the resistor?

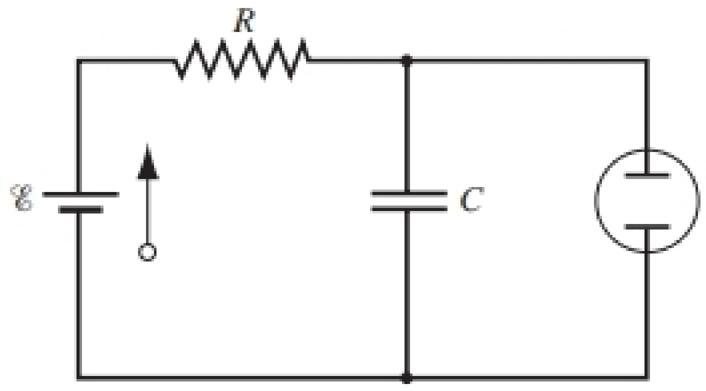


FIGURE 31-39. Exercise 47.

48. A 1.0-μF capacitor with an initial stored energy of 0.50 J is discharged through a 1.0-MΩ resistor. (*a*) What is the initial charge on the capacitor? (*b*) What is the current through the resistor when the discharge starts? (*c*) Determine ΔV_C, the voltage across the capacitor, and ΔV_R, the voltage across the resistor, as functions of time. (*d*) Express the rate of generation of internal energy in the resistor as a function of time.

49. A 3.0-MΩ resistor and a 1.0-μF capacitor are connected in a single-loop circuit to a source of emf with $\mathcal{E} = 4.0$ V. At 1.0 s after the connection is made, what are the rates at which (*a*) the charge on the capacitor is increasing, (*b*) energy is being stored in the capacitor, (*c*) internal energy is appearing in the resistor, and (*d*) energy is being delivered by the seat of emf?

50. Prove that when switch S in Fig. 31-20 is thrown from *a* to *b*, all the energy stored in the capacitor is transformed into internal energy in the resistor. Assume that the capacitor is fully charged before the switch is thrown.

PROBLEMS

1. The starting motor of an automobile is turning slowly and the mechanic has to decide whether to replace the motor, the cable, or the battery. The manufacturer's manual says that the 12-V battery can have no more than 0.020-Ω internal resistance, the motor no more than 0.200-Ω resistance, and the cable no more than 0.040-Ω resistance. The mechanic turns on the motor and measures 11.4 V across the battery, 3.0 V across the cable, and a current of 50 A. Which part is defective?

2. Two batteries having the same emf $\mathcal{E}$ but different internal resistances r_1 and r_2 ($r_1 > r_2$) are connected in series to an external resistance *R*. (*a*) Find the value of *R* that makes the potential difference zero between the terminals of one battery. (*b*) Which battery is it?

3. A solar cell generates a potential difference of 0.10 V when a 500-Ω resistor is connected across it and a potential difference of 0.16 V when a 1000-Ω resistor is substituted. What are (*a*) the internal resistance and (*b*) the emf of the solar cell? (*c*) The area of the cell is 5.0 cm^2 and the intensity of light striking it is 2.0 mW/cm^2. What is the efficiency of the cell for converting light energy to internal energy in the 1000-Ω external resistor?

4. When the lights of an automobile are switched on, an amme-

drops to 8.00 A and the lights dim somewhat. If the internal resistance of the battery is 50.0 mΩ and that of the ammeter is negligible, what are (*a*) the emf of the battery and (*b*) the current through the starting motor when the lights are on?

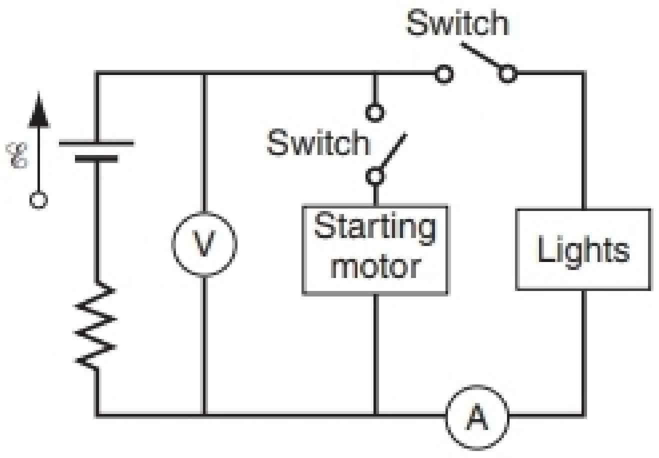


FIGURE 31-40. Problem 4.

5. Conducting rails *A* and *B*, having equal lengths of 42.6 m and cross-sectional area of 91.0 cm^2, are connected in series. A potential difference of 630 V is applied across the terminal points of the connected rails. The resistances of the rails are 76.2 $\mu\Omega$ and 35.0 $\mu\Omega$. Determine (*a*) the resistivities of the

14. A coil of current-carrying Nichrome wire is immersed in a liquid contained in a calorimeter. When the potential difference across the coil is 12 V and the current through the coil is 5.2 A, the liquid boils at a steady rate, evaporating at the rate of 21 mg/s. Calculate the heat of vaporization of the liquid.

15. A resistance coil, wired to an external battery, is placed inside an adiabatic cylinder fitted with a frictionless piston and containing an ideal gas. A current $i = 240$ mA flows through the coil, which has a resistance $R = 550\ \Omega$. At what speed v must the piston, mass $m = 11.8$ kg, move upward so that the temperature of the gas remains unchanged? See Fig. 31-46.

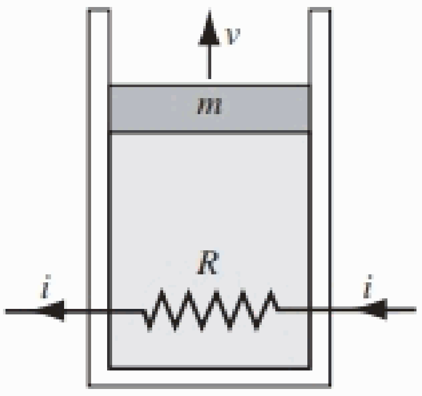


FIGURE 31-46. Problem 15.

16. A 32-μF capacitor is connected across a programmed power supply. During the interval from $t = 0$ to $t = 3$ s, the output voltage of the supply is given by $V(t) = (6\text{ V}) + (4\text{ V/s})t - (2\text{ V/s}^2)t^2$. At $t = 0.50$ s, find (*a*) the charge on the capacitor, (*b*) the current into the capacitor, and (*c*) the power output from the power supply.

17. A potential difference ΔV is applied to a wire of cross-sectional area A, length L, and conductivity σ. You want to change the applied potential difference and draw out the wire so the power dissipated is increased by a factor of 30 and the current is increased by a factor of 4. What should be the new values of the (*a*) length and (*b*) cross-sectional area?

18. An initially uncharged capacitor C is fully charged by a constant emf $\mathcal{E}$ in series with a resistor R. (*a*) Show that the final energy stored in the capacitor is half the energy supplied by the emf. (*b*) By direct integration of i^2R over the charging time, show that the internal energy dissipated by the resistor is also half the energy supplied by the emf.

19. At what time after charging begins in Problem 18 is the rate of energy dissipation in the resistor equal to the rate of energy storage in the capacitor?

COMPUTER PROBLEMS

1. Use the results of Exercise 28 to find the equivalent resistance between any two points separated by a single resistor (such as A and B) for the infinite resistor array in Fig. 31-37. This problem is easiest done by iteration, and can be programmed and solved on a spreadsheet in less than a minute!

2. Repeat Computer Problem 1 for a torus instead of an infinite sheet. Start with a torus made from a 10 × 10 grid of resistors. By how much does the answer change if the original grid is doubled in size? Refer to Computer Problem 2 of Chapter 30 for hints on setting up this problem.

CHAPTER 32

THE MAGNETIC FIELD

The science of magnetism *had its origin in ancient times. It grew from the observation that certain naturally occurring stones would attract one another and would also attract small bits of one metal, iron, but not other metals, such as gold or silver. The word "magnetism" comes from the name of the district (Magnesia) in Asia Minor, one of the locations where these stones were found.*

Today we have put that discovery to great practical use, from small "refrigerator" magnets to magnetic recording tape and computer disks. The magnetism of individual atomic nuclei is used by physicians to make images of organs deep within the body. Spacecraft have measured the magnetism of the Earth and the other planets to learn about their internal structure.

In this chapter we begin our study of magnetism by considering the magnetic field and its effects on a moving electric charge. In the next chapter we consider the production of magnetic fields by electric currents. In later chapters we continue to explore the close relationship between electricity and magnetism, which are linked together under the common designation electromagnetism.

32-1 MAGNETIC INTERACTIONS AND MAGNETIC POLES

When we began our study of electrostatics in Chapter 25, we described an observation—the attraction between a plastic rod rubbed with fur and a glass rod rubbed with silk—that we could not explain based on the forces and interactions we had considered up to that point. We accounted for that attraction as a force exerted by the electric charges of one rod on the electric charges of the other. In subsequent chapters we learned that many interesting and useful phenomena can be understood in terms of this basic electrostatic force.

In this chapter we introduce another new observation, which will prove to have equally interesting and useful consequences. This observation is based on the *magnetic* interaction between objects, the effects of which are discussed in the next several chapters. As we begin this study, it is important to keep in mind many important similarities between electric and magnetic interactions as well as the important differences between them.

As early as the 8th century B.C., the Greeks had discovered that a piece of the mineral magnetite (known as lodestone, an oxide of iron) can attract a piece of iron but does not exert a measurable force on most other materials. Later it was discovered that one piece of magnetite can either attract or repel another piece, depending on their relative orientation. In these experiments and the ones we describe below, it is necessary to determine that neither of the objects carries a net electric charge, so that we can be sure these new forces cannot be of electrical origin.

By the 12th century, the following experiment was known. A small needle-shaped piece of magnetite is suspended so that it can pivot about a vertical axis. Even with no other magnetite or iron nearby, the piece will spontaneously rotate about its pivot and eventually come to rest with one end pointing roughly toward the Earth's north geographic pole. Let us identify that end by painting it red. No

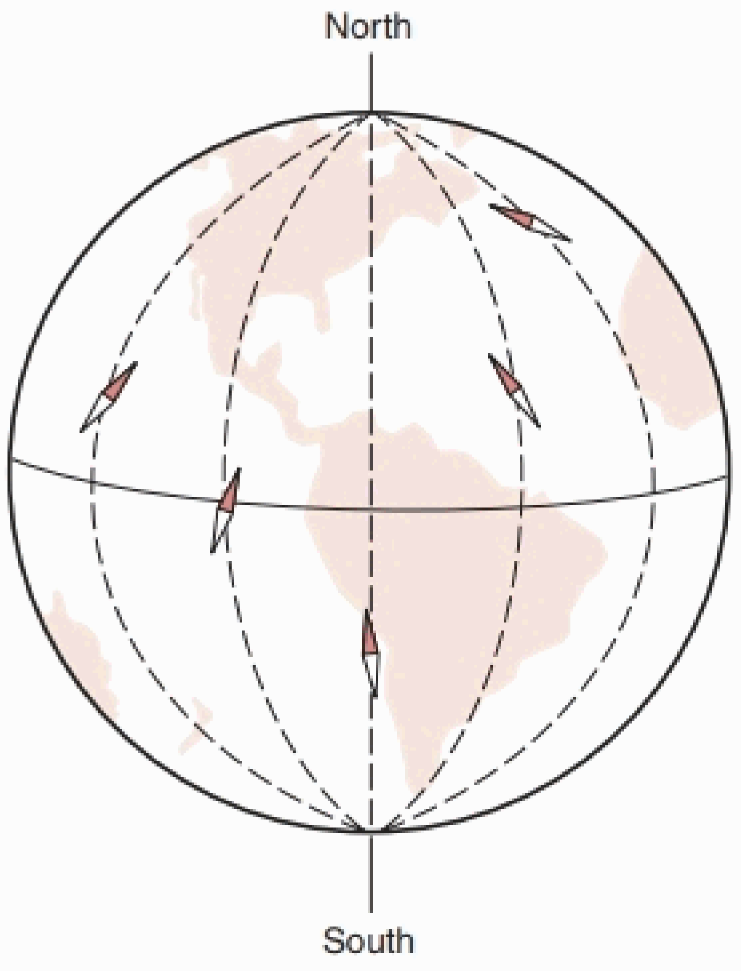


FIGURE 32-1. At any location on the Earth, a magnetic needle will point approximately in the direction of the north pole.

matter where we try this experiment on Earth, the red end always points north (Fig. 32-1).

The device we have constructed is of course a magnetic compass, which is responding to the magnetic field of the Earth, just as two pieces of magnetite exert forces on each other. We describe this outcome using the same language that we used for electrical interactions: one piece of magnetite (or the Earth) sets up a magnetic field, and the other piece responds to that magnetic field. The direction of a magnetic compass gives the direction of the horizontal component of the magnetic field of the Earth.

We can also use our magnetic compass to trace out the magnetic field of a bar magnet. In Fig. 32-2, one end of a bar magnet has been painted red, because that end points north if we suspend the bar magnet like a compass. By convention, we call this the "north-seeking" or simply *north* pole of the magnet, and the opposite end is the *south* pole. When we bring our compass close to the bar magnet, the compass rotates until its direction indicates the direction of the magnetic field at that point, as shown in Fig. 32-2. As we discuss in Chapter 35, the magnetic field of the Earth is in many ways similar to the magnetic field of a bar magnet.

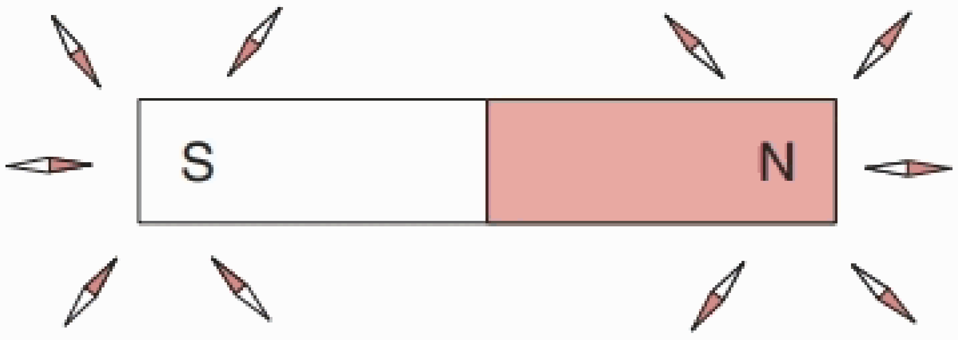


FIGURE 32-2. The compass needle shows the magnetic field surrounding a bar magnet.

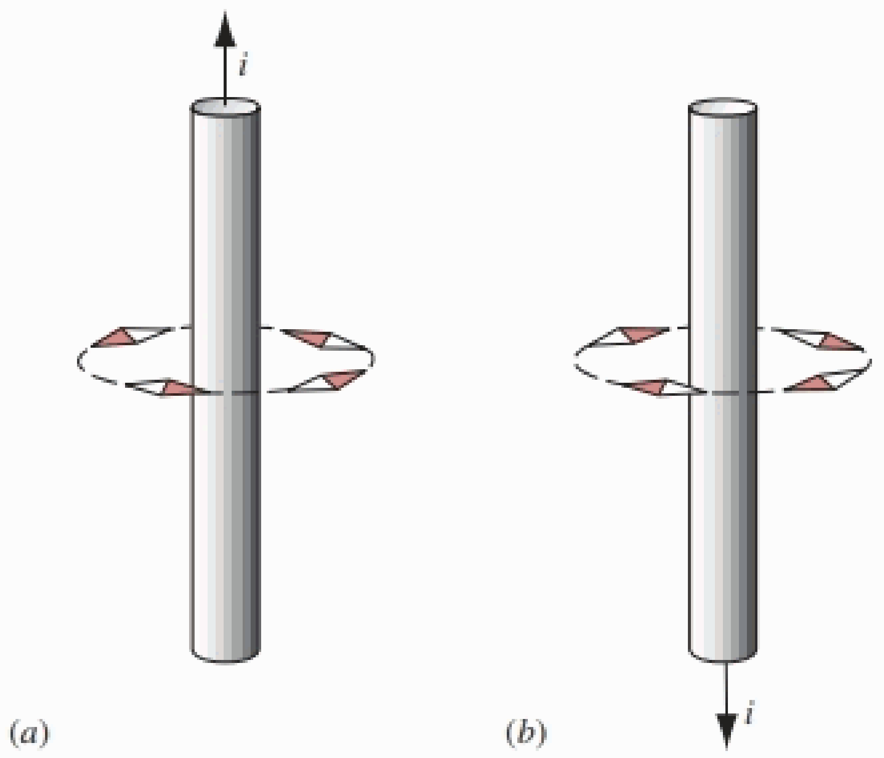


FIGURE 32-3. (*a*) A compass shows that a magnetic field surrounds a current-carrying wire. (*b*) When the current is reversed, the direction of the magnetic field reverses.

Even more surprising results occur when we place our compass near a current-carrying wire, as in Fig. 32-3. When a steady current flows in the wire, the compass shows quite clearly that a magnetic field is present, and the direction of the compass needle indicates the direction of the magnetic field near the wire. If we turn off the current, there is no magnetic field. If we reverse the direction of the current, the compass needle points in the opposite direction (Fig. 32-3*b*).

These observations give a hint of the complex and fascinating relationship between electrical phenomena (such as a current in a wire) and magnetic phenomena (the deflection of a compass needle). Later in the text we discuss other examples of this relationship, which is responsible for such diverse effects as the operation of electric motors and the propagation of light.

Magnetism and Moving Charges

It is tempting to try to understand magnetic fields by following the same procedure we used for electric fields—that is, by using a test charge to probe the field. This leads immediately to questions about the relationship between electric and magnetic phenomena.

1. *Does there exist in nature a "magnetic test charge" that could be used to determine the strength and direction of the magnetic field, just as we used the force on an electric test charge to determine the electric field?* (As in $\vec{\mathbf{E}} = \vec{\mathbf{F}}/q_0$, Eq. 26-3). Theory permits the existence of isolated magnetic charges, but no one has ever found one despite intense experimental searches. We conclude that isolated magnetic charges, known as *magnetic monopoles,* are either very rare or nonexistent, and so our equations for electromagnetism are written as if there are no magnetic charges.

2. *Can we use an electric test charge to probe a magnetic field?* Yes, but only if the charge is moving relative to the source of the magnetic field. A magnetic field exerts no force on an electric charge at rest.

3. *If electric charges in motion can be used to probe magnetic fields, can electric charges in motion also produce magnetic fields?* Yes, as we illustrated in Fig. 32-3. In fact, electric charges in motion are also responsible for the magnetic fields of the Earth and the bar magnet (in the latter case, the charges are electrons moving in atoms).

4. *In electrostatics we associated an electric potential energy with a test charge in an electric field (Section 28-2). Is there a "magnetic potential energy" associated with a moving electric test charge in a magnetic field?* In general the answer is no, because forces that depend on velocity are not conservative forces. (Recall from Chapter 12 that potential energy can be defined only for conservative forces.)

Just as the space around a collection of charges is described as the location of an electric field represented by a vector $\vec{\mathbf{E}}$, the space around the Earth, a bar magnet, or a current-carrying wire is described as the location of a magnetic field* represented by a vector $\vec{\mathbf{B}}$. In analogy with the electric field, we display the magnetic field by field lines, which are close together where the field is large and far apart where it is small. Often the lines of $\vec{\mathbf{B}}$ will look similar to patterns we have previously drawn for lines of $\vec{\mathbf{E}}$—for example, in the case of a uniform field or the field of a dipole. However, as we go along you should take note of several important differences between the lines of $\vec{\mathbf{E}}$ and the lines of $\vec{\mathbf{B}}$, which we shall discuss in the next several chapters.

The basic relationship between electric charge and electric field in electrostatics can be represented as

$$\overbrace{\text{electric charge} \rightarrow \rlap{$\underbrace{\phantom{\text{electric field} \rightarrow \text{electric charge.}}}_{\substack{\text{forces are exerted on electric}\\ \text{charges by electric fields}}}$}\text{electric field}}^{\substack{\text{electric fields are caused}\\ \text{by electric charge}}} \rightarrow \text{electric charge.}$$

That is, electric charges set up an electric field, which can then exert a force of electric origin on other electric charges. We would like to be able to write a similar relationship for magnetic fields:

$$\text{magnetic charge} \rightarrow \text{magnetic field} \rightarrow \text{magnetic charge.}$$

However, because no isolated magnetic charge has yet been found, we must use a different relationship:

$$\text{moving electric charge} \rightarrow \text{magnetic field} \rightarrow \text{moving electric charge.}$$

That is, a magnetic field is established by moving electric charges, and in turn the field can exert a force (which we call a magnetic force) on other moving electric charges. In this chapter we explore the second part of this relationship (the magnetic force on a moving charge). The next chapter discusses how magnetic fields are caused by moving electric charges, including currents in wires.

Magnetic Poles

By suspending a piece of magnetic material at the Earth's surface, we can identify and mark its north pole (the end that points roughly in the direction of the Earth's north geographic pole) and its south pole (the opposite end). Let us test and mark two pieces of magnetic material in this way. We can then directly study the magnetic force that one of these pieces exerts on the other in various orientations. In particular, we can study the force that one north pole exerts on another north pole or on a south pole. From experiments of this type we deduce the following rule for the interactions of magnetic poles:

Like poles repel, and unlike poles attract.

That is, a south pole attracts a north pole, but two north poles or two south poles repel one another. This rule is very similar to the rule for the interaction of electric charges (Section 25-2). Applying this rule to the behavior of a magnetic compass on the Earth's surface, as in Fig. 32-1, we conclude that to attract the north pole of the compass there must be a magnetic *south* pole near the Earth's geographic north pole.

32-2 THE MAGNETIC FORCE ON A MOVING CHARGE

To begin to understand the properties of magnetic fields, our first task is to study the force on a charged particle moving in a magnetic field. In electrostatics, the force on a charged particle in an electric field is $\vec{\mathbf{F}}_E = q\vec{\mathbf{E}}$ (Eq. 26-4). This expression gives us a way to test whether an electric field is present at various points in any region of space; if, after we account for all other nonelectric forces (gravity, etc.), we find the force on a test charge at rest to be nonzero, then we can conclude that an electric field must be present at that point. We now seek the corresponding expression for magnetic fields, which will allow us to test for the presence of magnetic fields based on the force exerted on a moving charged particle. We know in advance that we will not obtain as simple an expression as the electric force, because the magnetic force involves two vectors: the magnetic field $\vec{\mathbf{B}}$ and the velocity $\vec{\mathbf{v}}$.

Before beginning our experiments, we first test the region of interest to make sure that no electric field is present. We make this determination by placing the test charge at rest at various locations and checking that the electric force on the test charge is zero.

---

*There is not general agreement on the naming of field vectors in magnetism. $\vec{\mathbf{B}}$ may be called the *magnetic induction* or *magnetic flux density*, whereas another field vector, denoted by $\vec{\mathbf{H}}$, may be called the magnetic field. We regard $\vec{\mathbf{B}}$ as the more fundamental quantity and therefore call it the magnetic field.

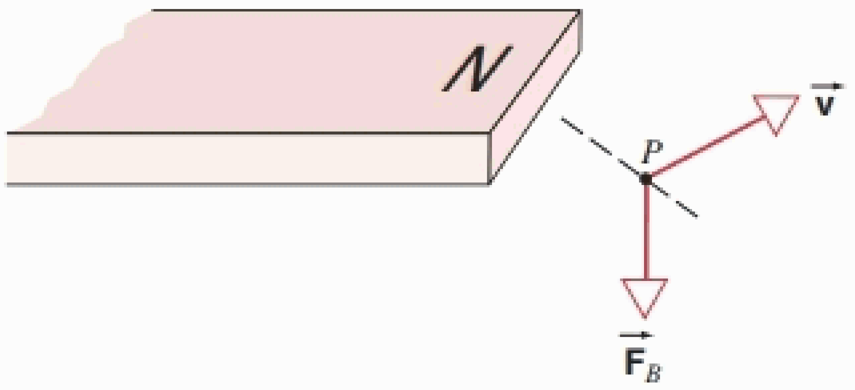


FIGURE 32-4. A charged particle moves with velocity $\vec{\mathbf{v}}$ near a bar magnet. A magnetic force $\vec{\mathbf{F}}_B$ is exerted on the particle. If the particle's velocity is in either direction along the dashed line, the magnetic force is zero.

Once we have established that there is no electric field acting on the charged particle, we can use it as a probe to find the magnetic force. We pick a particular point P near a source of a magnetic field, such as a bar magnet (Fig. 32-4). We assume that we have available a device for firing charged particles through point P; the device allows us to choose the magnitude of the velocity of the particles and their direction of motion. Using this device, we can do experiments that permit a variety of observations about the magnetic force $\vec{\mathbf{F}}_B$ that acts on the particles.

1. By firing particles through P in a variety of directions, we find that the magnetic force, if it is present, is perpendicular to the velocity $\vec{\mathbf{v}}$, as shown in Fig. 32-4. No matter what the direction of $\vec{\mathbf{v}}$, the magnetic force is always perpendicular to $\vec{\mathbf{v}}$.

2. As we vary the direction of $\vec{\mathbf{v}}$ through P (keeping the magnitude of $\vec{\mathbf{v}}$ constant), we find that the magnetic force is zero for one particular direction of $\vec{\mathbf{v}}$, indicated by the dashed line in Fig. 32-4. When the velocity is perpendicular to this direction, the magnetic force has its maximum value. In between these directions, the magnetic force varies as $\sin\phi$, where ϕ is the angle between the velocity and the dashed line. (Note that the dashed line actually indicates two opposite directions for which the force is zero, one corresponding to $\phi = 0$ and the other to $\phi = 180°$.)

3. As we vary the magnitude of the velocity, we find that the magnitude of $\vec{\mathbf{F}}_B$ varies in direct proportion.

4. We also find that the magnitude of $\vec{\mathbf{F}}_B$ is proportional to the magnitude of the test charge q, and that $\vec{\mathbf{F}}_B$ reverses direction when q changes sign.

We now define the magnetic field $\vec{\mathbf{B}}$ in the following way, based on these observations: the direction of $\vec{\mathbf{B}}$ at point P is the same as one of the directions of $\vec{\mathbf{v}}$ (to be specified shortly) for which the force is zero, and the magnitude of $\vec{\mathbf{B}}$ is determined from the magnitude $F_{B,\max}$ of the maximum force exerted when the test charge is projected perpendicular to the direction of $\vec{\mathbf{B}}$; that is,

$$B = \frac{F_{B,\max}}{|q|v}. \tag{32-1}$$

At arbitrary angles, our observations are summarized by the formula

$$F_B = |q|vB\sin\phi, \tag{32-2}$$

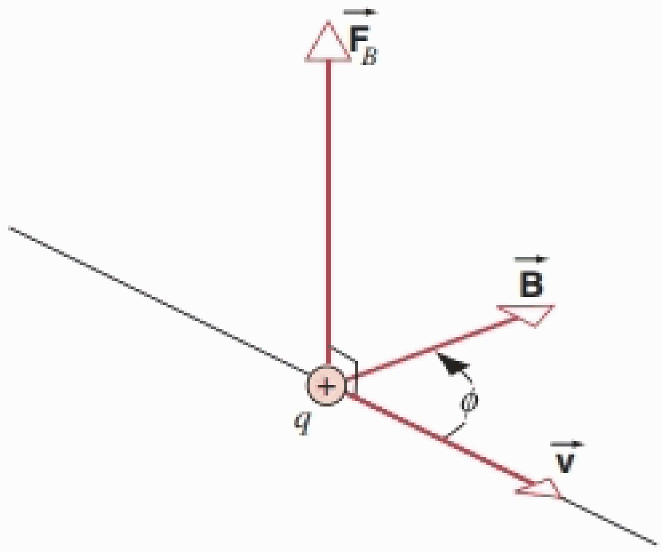


FIGURE 32-5. A particle with a positive charge q moving with velocity $\vec{\mathbf{v}}$ through a magnetic field $\vec{\mathbf{B}}$ experiences a magnetic deflecting force $\vec{\mathbf{F}}_B$.

where ϕ is the smaller angle between $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$. Because F_B, v, and B are vectors, Eq. 32-2 can be written as a vector product:

$$\vec{\mathbf{F}}_B = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}. \tag{32-3}$$

By writing $\vec{\mathbf{v}} \times \vec{\mathbf{B}}$ instead of $\vec{\mathbf{B}} \times \vec{\mathbf{v}}$ in Eq. 32-3, we have specified which of the two possible directions of $\vec{\mathbf{B}}$ we want to use. Vector or cross products will occur frequently in our study of magnetism. You may find it helpful to review the properties of the vector product in Appendix H.

Figure 32-5 shows the geometrical relationship among the vectors $\vec{\mathbf{F}}_B$, $\vec{\mathbf{v}}$, and $\vec{\mathbf{B}}$. Note that, as is always the case for a vector product, $\vec{\mathbf{F}}_B$ is perpendicular to the plane formed by $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$. Thus $\vec{\mathbf{F}}_B$ is always perpendicular to $\vec{\mathbf{v}}$, and the magnetic force is always a sideways deflecting force. Note also that $\vec{\mathbf{F}}_B$ vanishes when $\vec{\mathbf{v}}$ is either parallel or antiparallel to the direction of $\vec{\mathbf{B}}$ (in which case $\phi = 0°$ or $180°$, and $\vec{\mathbf{v}} \times \vec{\mathbf{B}} = 0$), and that $\vec{\mathbf{F}}_B$ has its maximum magnitude, equal to qvB, when $\vec{\mathbf{v}}$ is at right angles to $\vec{\mathbf{B}}$.

Because the magnetic force is always perpendicular to $\vec{\mathbf{v}}$, it cannot change the magnitude of $\vec{\mathbf{v}}$, only its direction. Equivalently, the force is always at right angles to the displacement of the particle and can do no work on it. Thus a constant magnetic field cannot change the kinetic energy of a moving charged particle. (In Chapter 34 we consider time-varying magnetic fields, which *can* change the kinetic energy of a particle. In this chapter we deal only with magnetic fields that do not vary with time.)

Equation 32-3, which serves as the definition of the magnetic field $\vec{\mathbf{B}}$, indicates both its magnitude and its direction. We define the electric field similarly through an equation, $\vec{\mathbf{F}}_E = q\vec{\mathbf{E}}$, so that by measuring the electric force we can determine the magnitude *and* direction of the electric field. Magnetic fields cannot be determined quite so simply with a single measurement. As Fig. 32-5 suggests, measuring $\vec{\mathbf{F}}_B$ for a single $\vec{\mathbf{v}}$ is not sufficient to determine $\vec{\mathbf{B}}$, because the direction of $\vec{\mathbf{F}}_B$ does not indicate the direction of $\vec{\mathbf{B}}$. We must first find the direction of $\vec{\mathbf{B}}$ (for example, by finding the directions of $\vec{\mathbf{v}}$ for which there is no force), and then a single additional measurement can determine its magnitude.

TABLE 32-1 Typical Values of Some Magnetic Fields[a]

Location	*Magnetic Field* (T)
At the surface of a neutron star (calculated)	10^8
Near a superconducting magnet	5
Near a large electromagnet	1
Near a small bar magnet	10^{-2}
At the surface of the Earth	10^{-4}
In interstellar space	10^{-10}
In a magnetically shielded room	10^{-14}

[a] Approximate values.

The SI unit of $\vec{\mathbf{B}}$ is the *tesla* (abbreviation T). It follows from Eq. 32-1 that

$$1 \text{ tesla} = 1 \frac{\text{newton}}{\text{coulomb} \cdot \text{meter/second}} = 1 \frac{\text{newton}}{\text{ampere} \cdot \text{meter}}.$$

An earlier (non-SI) unit for $\vec{\mathbf{B}}$, still in common use, is the *gauss,* related to the tesla by

$$1 \text{ tesla} = 10^4 \text{ gauss}.$$

Table 32-1 gives some typical values of magnetic fields.

Figure 32-6 shows the lines of $\vec{\mathbf{B}}$ of a bar magnet. Note that the lines of $\vec{\mathbf{B}}$ pass through the magnet, forming closed loops. From the clustering of field lines outside the magnet near its ends, we infer that the magnetic field in the space around the magnet has its greatest magnitude there. These ends are the *poles* of the magnet. Note that the field lines emerge from the north pole and converge toward the south pole.

SAMPLE PROBLEM 32-1. A uniform magnetic field $\vec{\mathbf{B}}$, with magnitude 1.2 mT, points vertically upward throughout the volume of the room in which you are sitting (Fig. 32-7). A proton with a kinetic energy of 5.3 MeV moves horizontally to the north through a certain point in the room. What magnetic deflecting force acts on the proton as it passes through this point? The proton mass is 1.67×10^{-27} kg.

Solution The magnetic deflecting force depends on the speed of the proton, which we can find from $K = \frac{1}{2}mv^2$. Solving for v, we find

$$v = \sqrt{\frac{2K}{m}} = \sqrt{\frac{(2)(5.3 \text{ MeV})(1.60 \times 10^{-13} \text{ J/MeV})}{1.67 \times 10^{-27} \text{ kg}}}$$
$$= 3.2 \times 10^7 \text{ m/s}.$$

Equation 32-2 then yields

$$\begin{aligned} F_B &= |q|vB \sin\phi \\ &= (1.60 \times 10^{-19} \text{ C})(3.2 \times 10^7 \text{ m/s})(1.2 \times 10^{-3} \text{ T})(\sin 90^\circ) \\ &= 6.1 \times 10^{-15} \text{ N}. \end{aligned}$$

This may seem like a small force, but it acts on a particle of small mass, producing a large acceleration; namely,

$$a = \frac{F_B}{m} = \frac{6.1 \times 10^{-15} \text{ N}}{1.67 \times 10^{-27} \text{ kg}} = 3.7 \times 10^{12} \text{ m/s}^2.$$

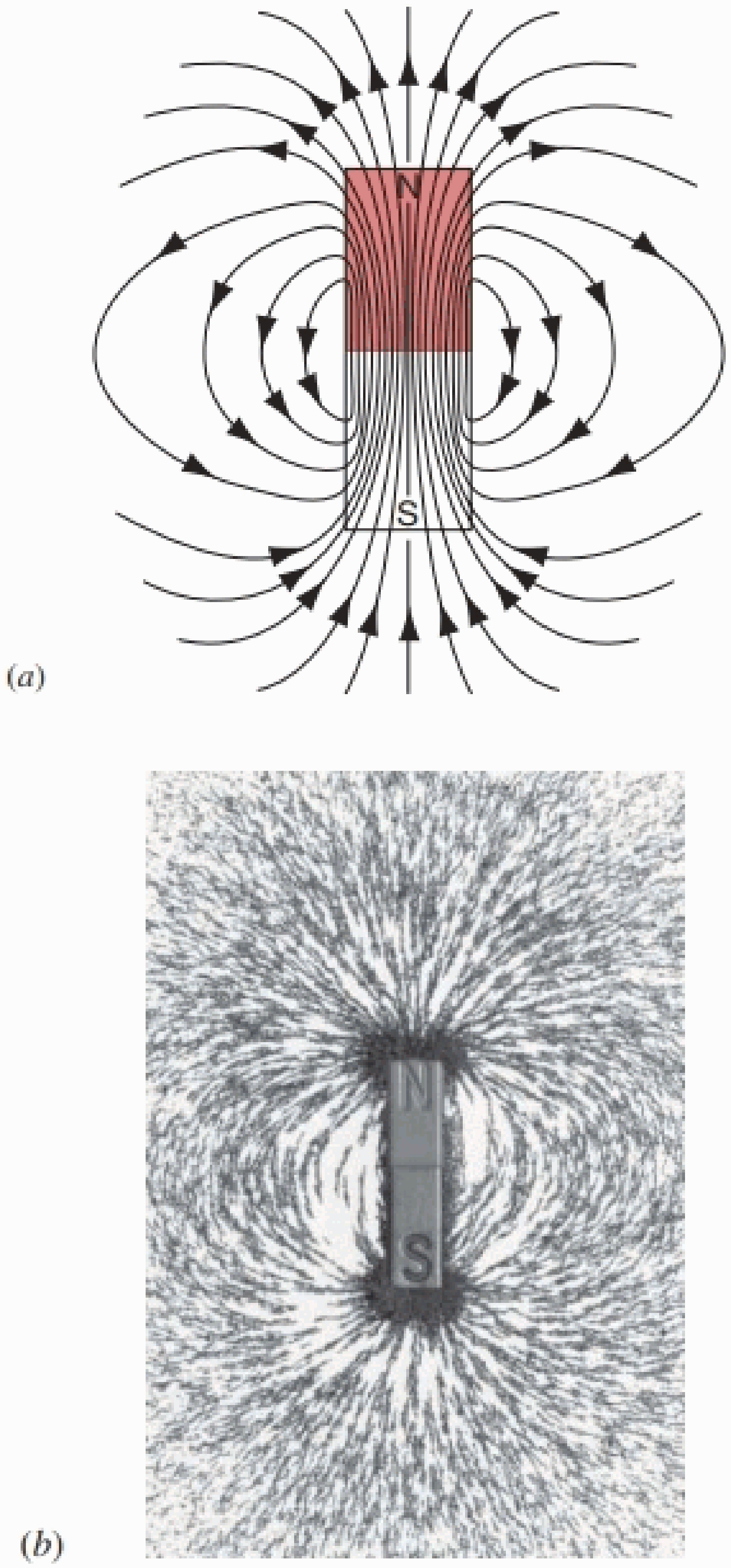


FIGURE 32-6. (*a*) The magnetic field lines for a bar magnet. The lines form closed loops, leaving the magnet at its north pole and entering at its south pole. (*b*) The field lines can be made visible by sprinkling iron filings on a sheet of paper covering a bar magnet.

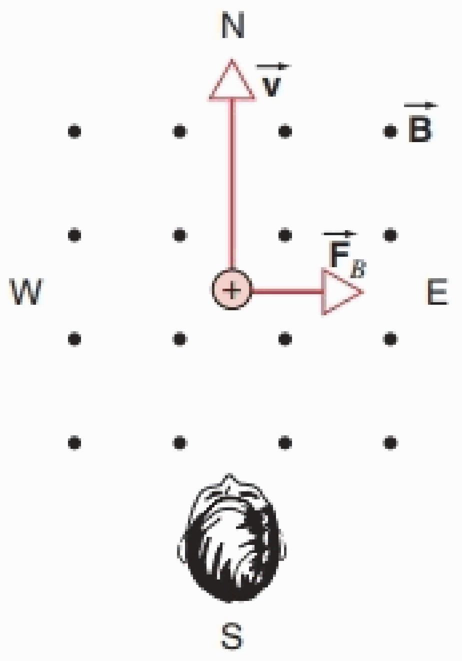


FIGURE 32-7. Sample Problem 32-1. A view from above of a student sitting in a room in which a vertically upward magnetic field deflects a moving proton toward the east. (The dots, which represent the points of arrows, symbolize magnetic field vectors pointing out of the page.)

It remains to find the direction of $\vec{F}_B$ when, as in Fig. 32-7, $\vec{v}$ points north, and $\vec{B}$ points vertically up. Using Eq. 32-3 and the right-hand rule for the direction of vector products (see Appendix H), we conclude that the deflecting force $\vec{F}_B$ must point horizontally to the east, as Fig. 32-7 shows.

If the charge of the particle had been negative, the magnetic deflecting force would have pointed in the opposite direction—that is, horizontally to the west. This is predicted automatically by Eq. 32-3, if we substitute $-q$ for q.

In this calculation, we used the (approximate) classical expression ($K = \frac{1}{2}mv^2$) for the kinetic energy of the proton rather than the (exact) relativistic expression (see Eq. 20-27). The criterion for safely using the classical expression is $K \ll mc^2$, where mc^2 is the rest energy of the particle. In this case $K = 5.3$ MeV, and the rest energy of a proton (see Appendix F) is 938 MeV. This proton passes the test, and we were justified in using the classical $K = \frac{1}{2}mv^2$ formula for the kinetic energy. In dealing with energetic particles, we must always be alert to this point.

Combined Electric and Magnetic Fields

If both an electric field $\vec{E}$ and a magnetic field $\vec{B}$ act on a charged particle, the total force on it can be expressed as

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}. \tag{32-4}$$

This force is called the *Lorentz force*. The Lorentz force is not a new kind of force; it is merely the sum of the electric and magnetic forces that may simultaneously act on a charged particle. The electric part of this force acts on any charged particle, whether at rest or in motion; the magnetic part acts only on moving charged particles.

One common application of the Lorentz force occurs when a beam of charged particles passes through a region in which the $\vec{E}$ and $\vec{B}$ fields are perpendicular to each other and to the velocity of the particles. If $\vec{E}$, $\vec{B}$, and $\vec{v}$ are oriented as shown in Fig. 32-8, then the electric force $\vec{F}_E = q\vec{E}$ is in the opposite direction to the magnetic force $\vec{F}_B = q\vec{v} \times \vec{B}$. We can adjust the magnetic and electric fields until the magnitudes of the forces are equal, in which case the Lorentz force is zero. In scalar terms,

$$qE = qvB \tag{32-5}$$

or

$$v = \frac{E}{B}. \tag{32-6}$$

The crossed $\vec{E}$ and $\vec{B}$ fields therefore serve as a *velocity selector:* only particles with speed $v = E/B$ pass through the region unaffected by the two fields, whereas particles with other speeds are deflected. This value of v is independent of the charge or mass of the particles.

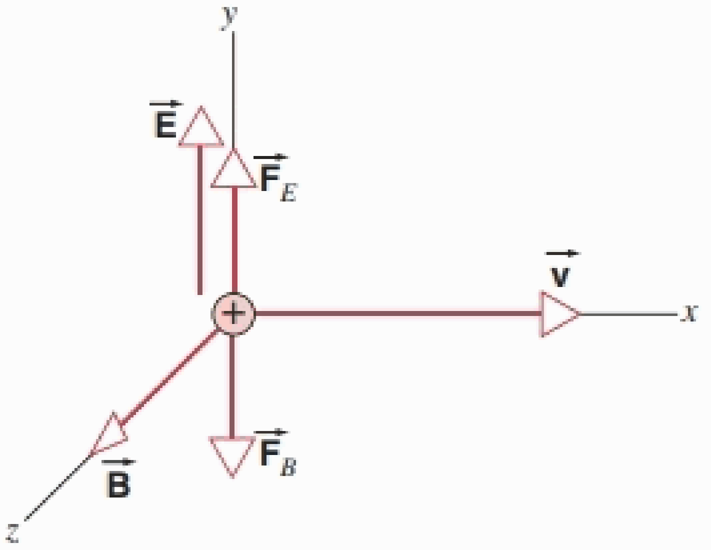


FIGURE 32-8. A positively charged particle, moving through a region in which there are electric and magnetic fields perpendicular to one another, experiences opposite electric and magnetic forces $\vec{F}_E$ and $\vec{F}_B$.

Beams of charged particles are often prepared using methods that give a distribution of speeds (for example, a thermal distribution such as that of Fig. 22-6). Using a velocity selector we can isolate particles with a chosen speed from the beam. This principle was applied in 1897 by J. J. Thomson in his discovery of the electron and the measurement of its charge-to-mass ratio. Figure 32-9 shows a modern version of his apparatus. Thomson first measured the vertical deflection y of the beam when only the electric field was present. From Sample Problem 26-6, the deflection is

$$y = -\frac{qEL^2}{2mv^2}. \tag{32-7}$$

In this expression, as in Fig. 32-9, we take the positive y direction to be upward, and E is the *magnitude* of the electric field. The deflection y of a negatively charged particle is positive in Eq. 32-7 and Fig. 32-9.

Then the magnetic field was turned on and adjusted until the beam deflection was zero (equivalent to that measured with no fields present). In this case $v = E/B$, and solving for the charge-to-mass ratio with $q = -e$ gives

$$\frac{e}{m} = \frac{2yE}{B^2L^2}. \tag{32-8}$$

Thomson's value for e/m (expressed in modern units) was 1.7×10^{11} C/kg, in good agreement with the current value of $1.758820174 \times 10^{11}$ C/kg.

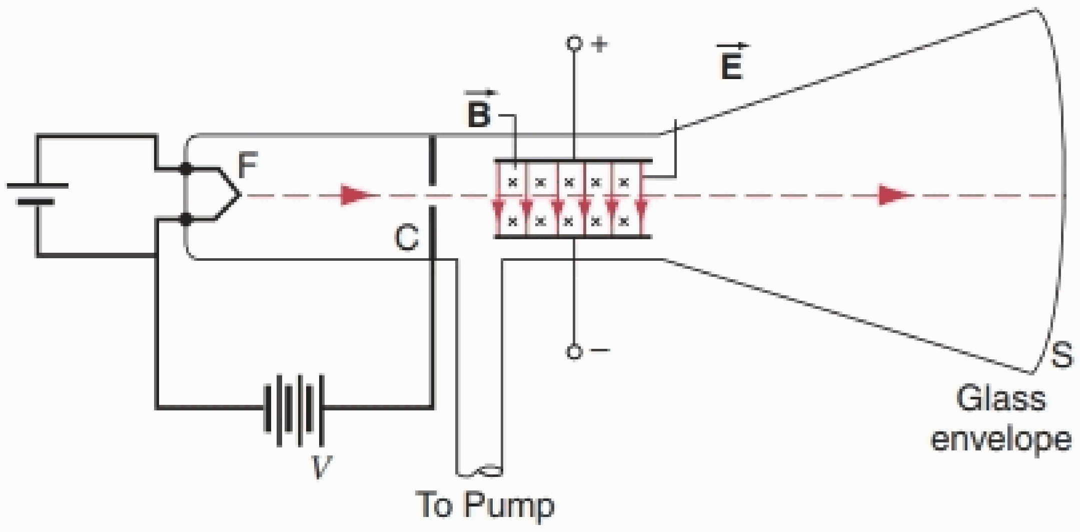


FIGURE 32-9. A modern version of J. J. Thomson's apparatus for measuring the charge-to-mass ratio of the electron. The filament F produces a beam of electrons with a distribution of speeds. The electric field $\vec{E}$ is set up by connecting a battery across the plate terminals. The magnetic field $\vec{B}$ is set up by means of current-carrying coils (not shown). The beam makes a visible spot where it strikes the screen S. (The crosses, which represent the tails of arrows, symbolize $\vec{B}$ vectors pointing into the page.)

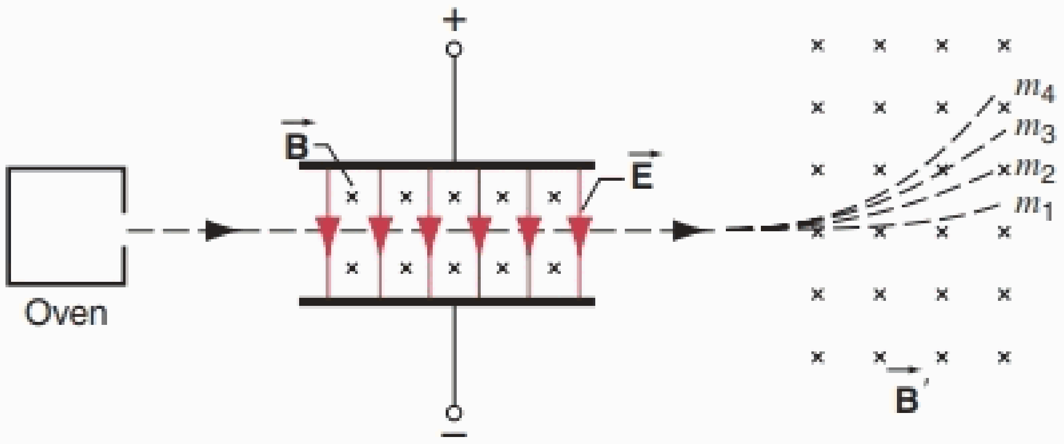


FIGURE 32-10. Schematic diagram of a mass spectrometer. A beam of ionized atoms having a mixture of different masses leaves an oven and enters a region of perpendicular $\vec{E}$ and $\vec{B}$ fields. Only those atoms with speeds $v = E/B$ pass undeflected through the region. Another magnetic field $\vec{B}'$ deflects the atoms along circular paths whose radii are determined by the masses of the atoms.

Another application of the velocity selector is in the mass spectrometer, a device for separating ions by mass. In this case a beam of ions, perhaps including species of differing masses, may be obtained from a vapor of the material heated in an oven (see Fig. 32-10). A velocity selector passes only ions of a particular speed, and when the resulting beam is then passed through another magnetic field, the paths of the particles are circular arcs (as we show in the next section) whose radii are determined by the momentum of the particles. Since all the particles have the same speed, the radius of the path is determined by the mass, and each different mass component in the beam follows a path of a different radius. These atoms can be collected and measured or else formed into a beam for subsequent experiments. See Problems 5 through 8 for other details on separating ions by their mass.

32-3 CIRCULATING CHARGES

Figure 32-11 shows a beam of electrons traveling through an evacuated chamber in which there is a uniform magnetic field $\vec{B}$. The magnetic deflecting force is the only important force that acts on the electrons. The beam clearly follows a circular path. Let us see how we can understand this behavior.

The magnetic deflecting force has two properties that affect the trajectories of charged particles: (1) it does not change the speed of the particles, and (2) it always acts perpendicular to the velocity of the particles. These are exactly the characteristics we require for a particle to move in a circle at constant speed, as in the case of the electrons in Fig. 32-11.

Since $\vec{B}$ is perpendicular to $\vec{v}$, the magnitude of the magnetic force can be written $|q|vB$, and Newton's second law with a centripetal acceleration of v^2/r gives

$$|q|vB = m\frac{v^2}{r} \tag{32-9}$$

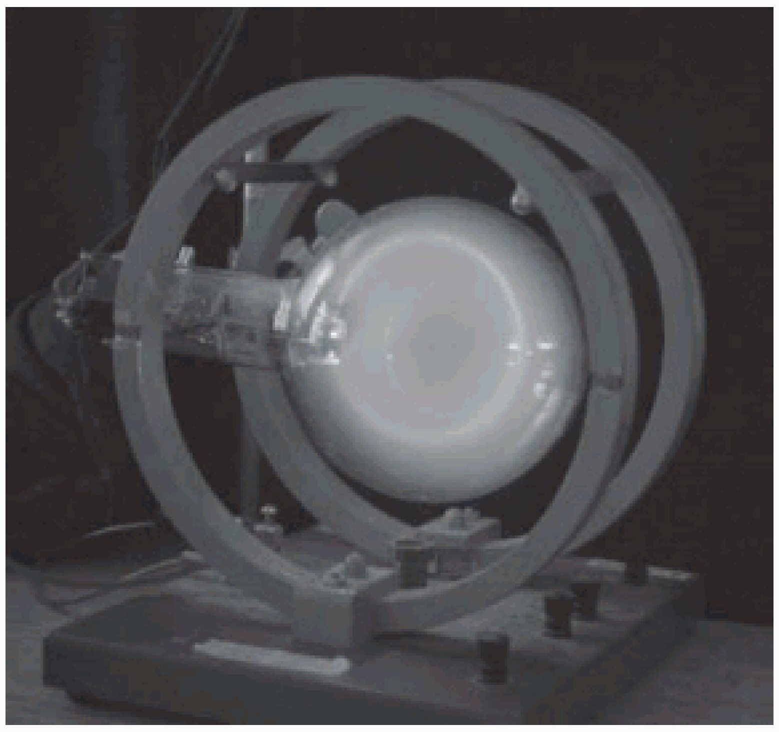


FIGURE 32-11. A beam of electrons travels in a circular path in a uniform magnetic field. The beam is made visible due to collisions with atoms of gas in the glass tube. The magnetic field is produced by the large coils; the direction of the field is along the axis of the coils. The electrons move in a plane perpendicular to the direction of the field. The magnetic force $\vec{F}_B$ on the electrons is directed radially inward.

or

$$r = \frac{mv}{|q|B} = \frac{p}{|q|B}. \tag{32-10}$$

Thus the radius of the path is determined by the momentum p of the particles, their charge, and the strength of the magnetic field. If the source of the electrons in Fig. 32-11 had projected them with a smaller speed, they would have moved in a circle of smaller radius.

The angular velocity of the circular motion is

$$\omega = \frac{v}{r} = \frac{|q|B}{m}, \tag{32-11}$$

and the corresponding frequency is

$$f = \frac{\omega}{2\pi} = \frac{|q|B}{2\pi m}. \tag{32-12}$$

Note that the frequency associated with the circular motion does not depend on the speed of the particle (as long as $v \ll c$, as we discuss below). Thus, if electrons in Fig. 32-11 were projected with a smaller speed, they would require the same time to complete the smaller circle that the faster electrons require to complete the larger circle. The frequency given by Eq. 32-12 is called the *cyclotron frequency,* because particles circulate at this frequency in a type of particle accelerator called a *cyclotron.* The frequency is characteristic of a particular particle moving in a particular magnetic field, just as the oscillating pendulum or the mass–spring system has its characteristic frequency.

The Cyclotron

A cyclotron (Fig. 32-12) is a device that accelerates beams of charged particles, which might be used in nuclear reaction experiments or for medical purposes. Figure 32-13 shows a schematic view of a cyclotron. It consists of two

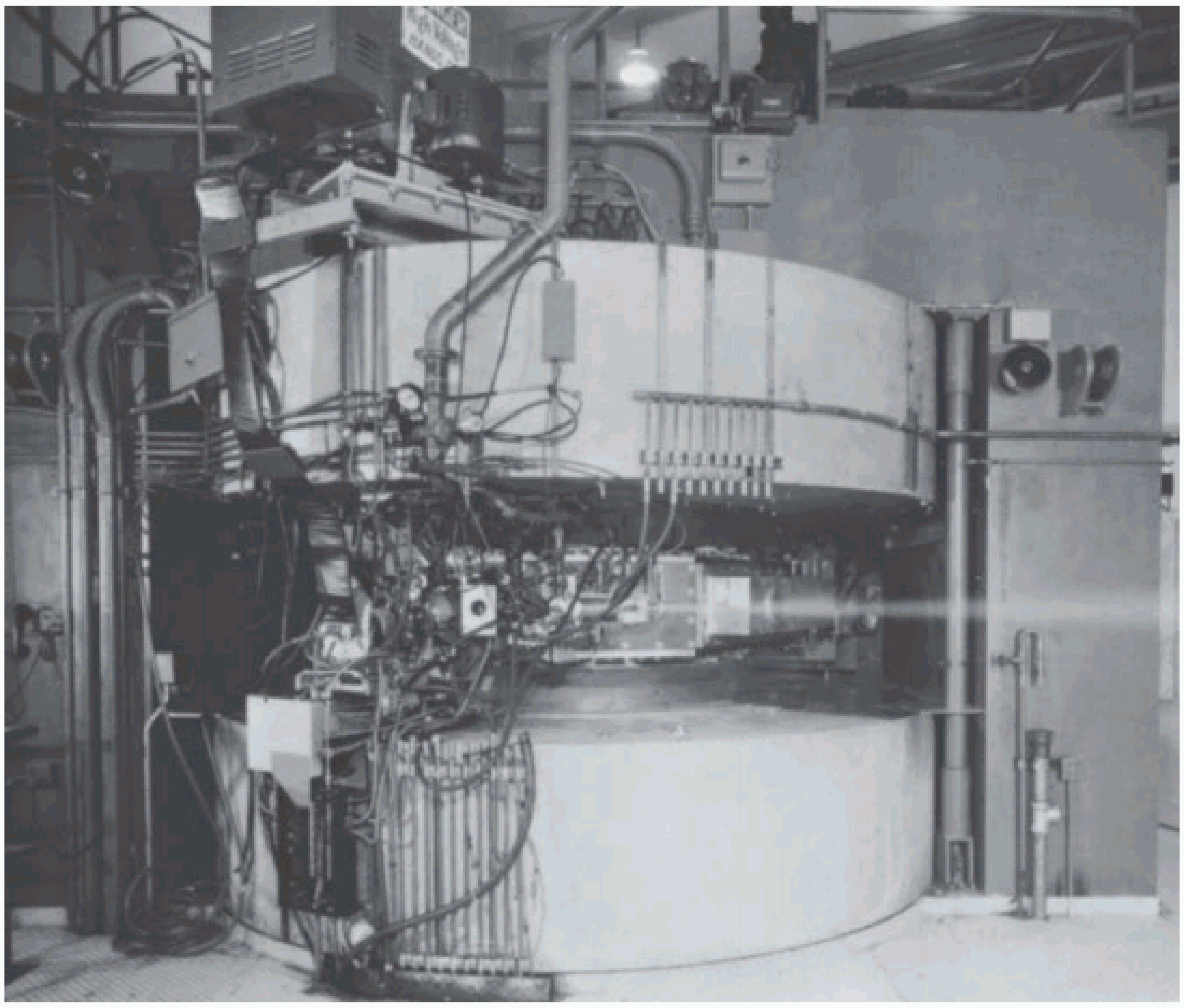


FIGURE 32-12. A cyclotron accelerator. The magnets are in the large chambers at the top and bottom. The beam is visible as it emerges from the accelerator because, like the beam of electrons in Fig. 32-11, it ionizes air molecules in collisions.

hollow metal D-shaped objects called *dees.* The dees are made of conducting material such as sheets of copper and are open along their straight edges. They are connected to an electric oscillator, which establishes an oscillating potential difference between the dees. A magnetic field is perpendicular to the plane of the dees. At the center of the instrument is a source that emits the ions we wish to accelerate.

When the ions are in the gap between the dees, they are accelerated by the potential difference between the dees. They then enter one of the dees, where they feel no electric field (because the electric field inside a conductor is zero), but the magnetic field (which is not shielded by the copper dees) bends their path into a semicircle. When the particles next enter the gap, the oscillator has reversed the direction of the electric field, and the particles are again accelerated as they cross the gap. Moving with greater speed, they travel a path of greater radius, as required by Eq. 32-10. However, according to Eq. 32-12, *it takes them exactly the same amount of time to travel the larger semicircle;* this is the critical characteristic of the operation of the cyclotron. The frequency of the electric oscillator must be adjusted to be equal to the cyclotron frequency (determined by the magnetic field and the charge and mass of the particle to be accelerated according to Eq. 32-12); this equality of frequencies is called the *resonance condition.* If the resonance condition is satisfied, particles continue to accelerate in the gap and "coast" around the semicircles, gaining a small increment of energy in each circuit, until they are deflected out of the accelerator.

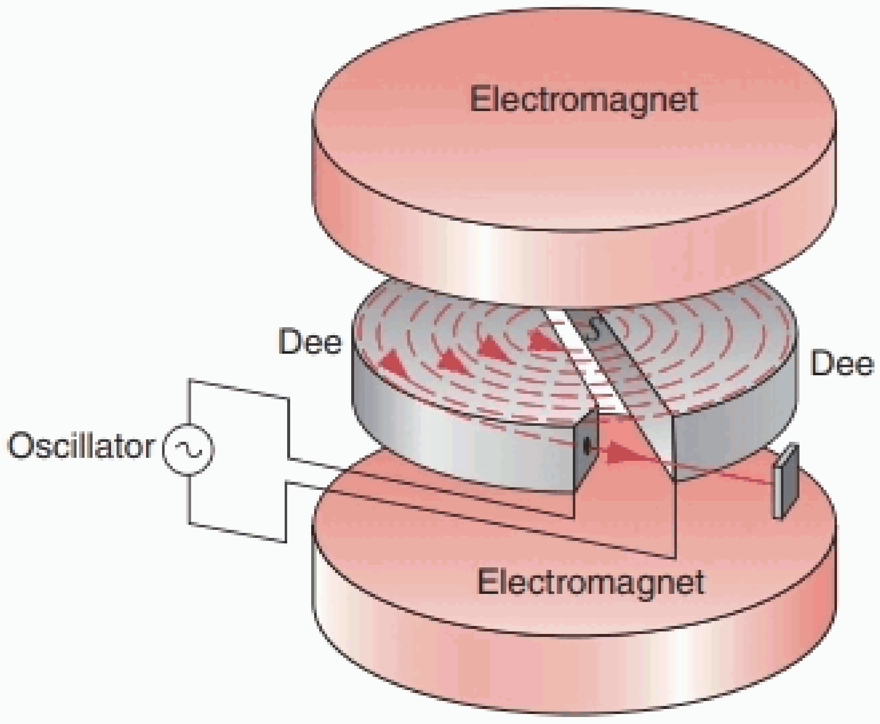


FIGURE 32-13. The elements of a cyclotron, showing the ion source S and the dees. The electromagnets provide a uniform vertical magnetic field. The particles spiral outward within the hollow dees, gaining energy every time they cross the gap between the dees.

The final speed of the particles is determined by the radius R at which the particles leave the accelerator. From Eq. 32-10,

$$v = \frac{|q|BR}{m}, \tag{32-13}$$

and the corresponding (nonrelativistic) kinetic energy of the particles is

$$K = \tfrac{1}{2}mv^2 = \frac{q^2B^2R^2}{2m}. \tag{32-14}$$

Typical cyclotrons produce beams of protons with maximum energies in the range of 10 MeV. For a given mass, ions with higher electric charges emerge with energies that increase with the square of the charge.

It is somewhat surprising that the energy in Eq. 32-14 depends on the magnetic field, which does not accelerate the particles, but does not depend on the electric potential difference that causes the acceleration. A larger potential difference gives the particles a larger "kick" with each cycle; the radius increases more quickly, and the particles make fewer cycles before leaving the accelerator. With a smaller potential difference, the particles make more cycles but get a smaller "kick" each time. Thus the energy of the particles is independent of the potential difference.

The cyclotron is limited in the energy to which it can accelerate particles, because as the particles reach greater speeds the classical expressions for momentum ($p = mv$ in Eq. 32-10) and kinetic energy ($K = \frac{1}{2}mv^2$ in Eq. 32-14) are no longer valid, and we must use the corresponding relativistic expressions from Chapter 20. As the particles begin to move at speeds that approach the speed of light, they take a longer time to travel the circular path and so the resonance condition is lost. In practical terms, about 40 MeV is the highest proton kinetic energy that can be achieved with a conventional cyclotron, which might have a radius of the order of 1 m.

To achieve the higher energies that are needed for current research in particle physics, an accelerator of a different design, called a *synchrotron,* is used. Even though the synchrotron does not rely on the cyclotron resonance condition, it uses magnetic fields to keep particles moving in a circular path in which they can be repeatedly accelerated by an electric field. In the design of a synchrotron, both the oscillator frequency and the strength of the magnetic field are varied as bunches of particles are accelerated in a circular path of constant radius. To achieve an energy of 1 TeV ($= 10^6$ MeV), as in the case of the Fermi National Accelerator Laboratory near Chicago, a path radius of 1 km is required; the protons make about 400,000 revolutions around the 6.3-km circumference in about 10 s.

The Magnetic Mirror

A nonuniform magnetic field can be used to trap a charged particle in a region of space. Figure 32-14 shows a schematic view of the operation of such a magnetic mirror. The charged particles tend to move in circles about the field direction. Suppose they also are drifting laterally, say, to the right in Fig. 32-14. The motion is therefore that of a helix, like a coiled spring. The field increases near the ends of the "magnetic bottle," and the force has a small component pointing toward the center of the region, which reverses the direction of the motion of the particles and causes them to spiral in the opposite direction, until they are eventually reflected from the opposite end. The particles continue to travel back and forth, confined to the space between the two high-field regions. Such an arrangement is used to confine the hot ionized gases (called *plasmas*) that are used in research into controlled thermonuclear fusion (see Section 51-8).

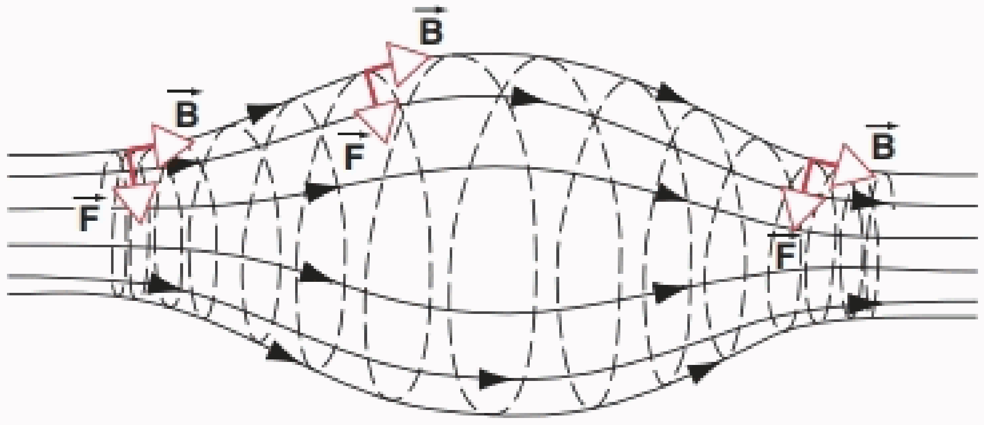


FIGURE 32-14. A charged particle spiraling in a nonuniform magnetic field. The field is greater at the left and right ends of the region than it is at the center. Particles can be trapped, spiraling back and forth between the strong-field regions at the ends. Note that the magnetic force vectors at each end of this "magnetic bottle" have components that point toward the center; it is these force components that serve to confine the particles.

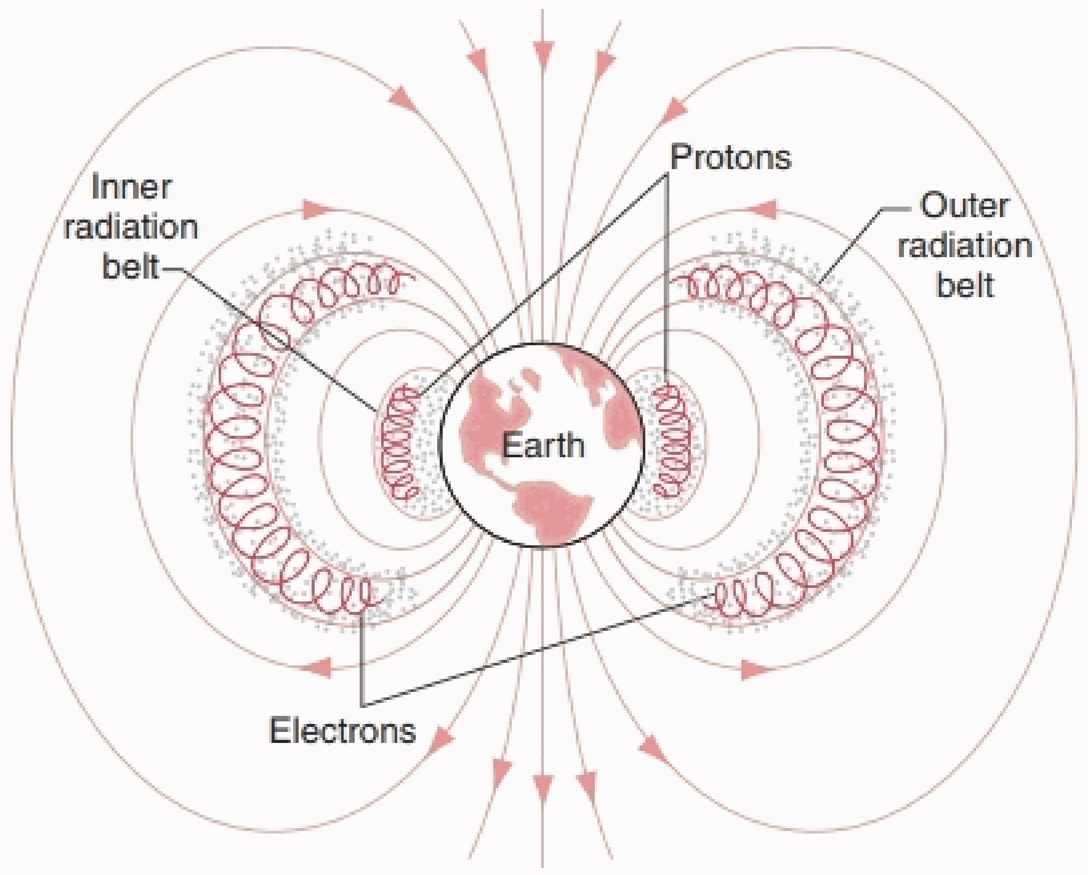


FIGURE 32-15. The Earth's magnetic field, showing protons and electrons trapped in the Van Allen radiation belts.

A similar phenomenon occurs in the Earth's magnetic field, as shown in Fig. 32-15. Electrons and protons are trapped in different regions of the Earth's field and spiral back and forth between the high-field regions near the poles in a time of a few seconds. These fast particles are responsible for the so-called Van Allen radiation belts that surround the Earth.

SAMPLE PROBLEM 32-2. A particular cyclotron is designed with dees of radius $R = 75$ cm and with magnets that can provide a field of 1.5 T. (*a*) To what frequency should the oscillator be set if deuterons are to be accelerated? (*b*) What is the maximum energy of deuterons that can be obtained?

Solution (*a*) A deuteron is a nucleus of heavy hydrogen, with a charge $q = +e$ and a mass of 3.34×10^{-27} kg, about twice the mass of ordinary hydrogen. Using Eq. 32-12 we can find the frequency:

$$f = \frac{|q|B}{2\pi m} = \frac{(1.60 \times 10^{-19}\ \mathrm{C})(1.5\ \mathrm{T})}{2\pi(3.34 \times 10^{-27}\ \mathrm{kg})}$$
$$= 1.1 \times 10^7\ \mathrm{Hz} = 11\ \mathrm{MHz}.$$

(*b*) The maximum energy occurs for deuterons that emerge at the maximum radius R. According to Eq. 32-14,

$$K = \frac{q^2B^2R^2}{2m} = \frac{(1.60 \times 10^{-19}\ \mathrm{C})^2(1.5\ \mathrm{T})^2(0.75\ \mathrm{m})^2}{2(3.34 \times 10^{-27}\ \mathrm{kg})}$$

$$= 4.85 \times 10^{-12}\ \mathrm{J} = 30\ \mathrm{MeV}.$$

Deuterons of this energy have a range in air of a few meters, as suggested by Fig. 32-12.

Motion in Nonuniform Fields (Optional)

The path of a charged particle in a nonuniform magnetic field is not a circle, and the calculation of the path (such as that shown in Fig. 32-14) can be a complicated problem. Usually there is no analytical solution, and the path must be calculated numerically, in analogy with the path of a projectile including air resistance (Chapter 4). Figure 32-16 illustrates why the path in a nonuniform field is no longer a circle. The particle is initially at point P_0, where its velocity $\vec{\mathbf{v}}_0$ is along the y direction. We assume that the field is in the z direction and that its strength increases as x and y increase. Initially the field is $\vec{\mathbf{B}}_0$, and the resulting force on the particle is $\vec{\mathbf{F}}_0$ in the x direction. This force gives the particle a velocity increment in the x direction, and an instant later the particle is at point P_1 moving with velocity $\vec{\mathbf{v}}_1$. The magnetic field $\vec{\mathbf{B}}_1$ at this point has a larger magnitude, and the force $\vec{\mathbf{F}}_1$ is also larger. For circular motion, the force must have the same magnitude at all locations, so the particle's path is clearly not a circle.

We can solve this problem numerically if we know how $\vec{\mathbf{B}}$ varies in magnitude (and possibly in direction) at all points where the particle might be located. We begin by writing the complete expression for the force as a vector (cross) product (see Appendix H):

$$\begin{aligned}\vec{\mathbf{F}}_B &= q(\vec{\mathbf{v}} \times \vec{\mathbf{B}}) \\ &= q(v_yB_z - v_zB_y)\hat{\mathbf{i}} + q(v_zB_x - v_xB_z)\hat{\mathbf{j}} \\ &\quad + q(v_xB_y - v_yB_x)\hat{\mathbf{k}}. \qquad (32\text{-}15)\end{aligned}$$

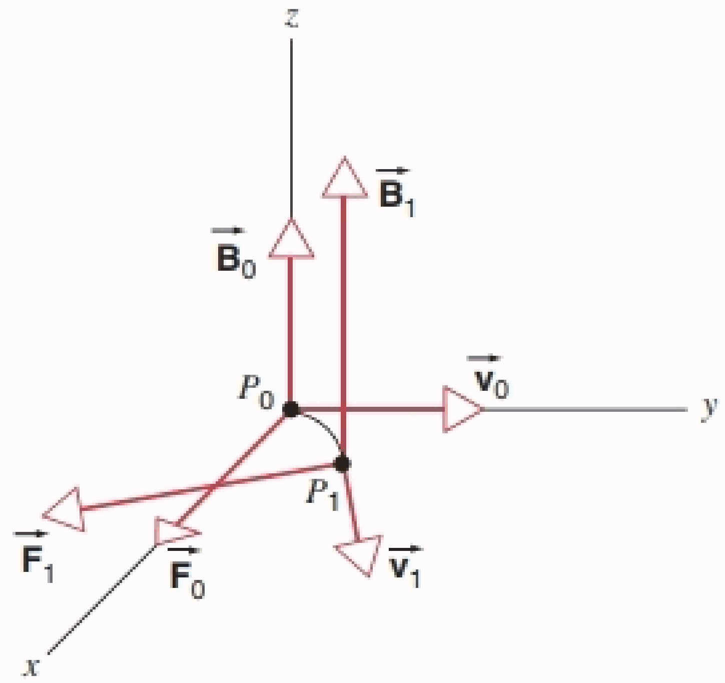


FIGURE 32-16. In a nonuniform magnetic field, the magnitude of the force varies (here as the particle moves from P_0 to P_1), and so the resulting path is not a circle.

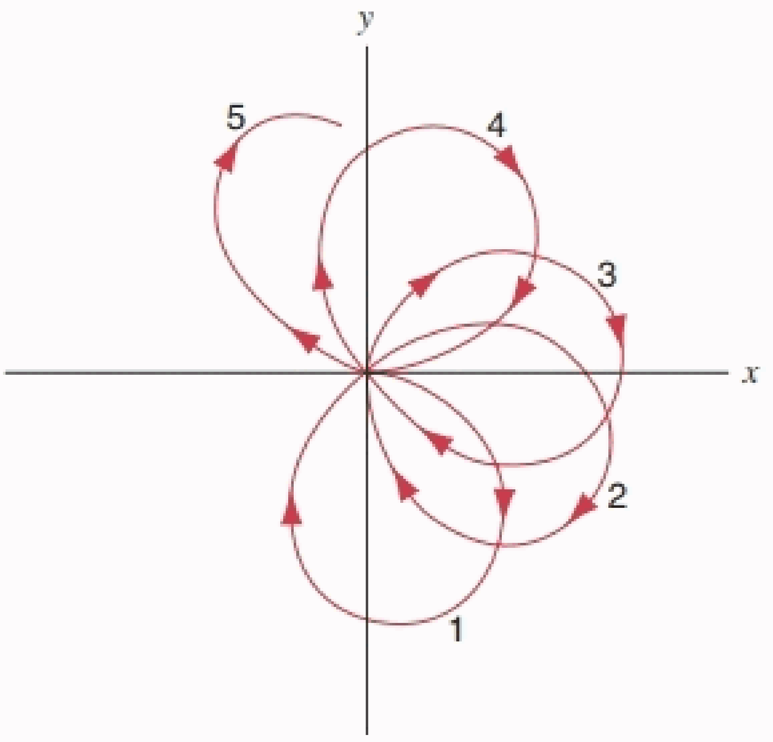


FIGURE 32-17. The path in the xy plane of a particle initially at the origin and moving in the x direction, subject to the magnetic field given by Eq. 32-17. The loops are numbered in the order they are traveled by the particle.

Using Newton's second law, $\vec{\mathbf{F}} = m\vec{\mathbf{a}} = m(a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}})$, we can equate the corresponding vector components to obtain the equations of motion, which can be solved for the path. For example, in the situation shown in Fig. 32-16, $\vec{\mathbf{B}}$ has only a z component ($B_x = B_y = 0$). In this case the equations of motion simplify to

$$\begin{aligned}F_{Bx} &= q(v_yB_z - v_zB_y) = qv_yB_z = ma_x = m\,dv_x/dt \\ F_{By} &= q(v_zB_x - v_xB_z) = -qv_xB_z = ma_y = m\,dv_y/dt \\ F_{Bz} &= q(v_xB_y - v_yB_x) = 0 = ma_z = m\,dv_z/dt. \qquad (32\text{-}16)\end{aligned}$$

(If $v_z = 0$ initially, then $v_z = 0$ at all times, because $a_z = 0$.) These expressions must be evaluated at every point of the trajectory, because the values of v_x, v_y, and B_z vary from point to point. Usually a computer is used to obtain a numerical solution for the path. For example, suppose the field varies as

$$B_z = B_0\left(1 + \frac{\sqrt{x^2 + y^2}}{R}\right), \qquad (32\text{-}17)$$

where $R = mv/qB_0$ is the radius of the path in a uniform field B_0. This field increases with the distance of the particle from the origin. The resulting motion, which can be obtained numerically (see Computer Problem 1), gives the flower-shaped pattern illustrated in Fig. 32-17. ■

32-4 THE HALL EFFECT

We have discussed several examples of the sideways deflecting force exerted on moving charged particles by a magnetic field. So far we have considered only individual particles or beams of particles moving otherwise freely. In 1879, Edwin Hall showed that the moving conduction electrons in a conductor can also be deflected by a magnetic field. The *Hall effect* provides a way to determine both the sign and the density of the charge carriers.

Consider a flat strip of material of width w carrying a current i, as shown in Fig. 32-18. The direction of the current i is the conventional one, opposite to the direction of motion of the electrons. A uniform magnetic field $\vec{\mathbf{B}}$ is established perpendicular to the plane of the strip, such as by placing the strip between the poles of an electromagnet. The charge carriers (electrons, for instance) experience a magnetic deflecting force $\vec{\mathbf{F}}_B = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}$, as shown in the figure, and move to the right side of the strip. Note that positive charges moving in the direction of i experience a deflecting force in the *same* direction.

The buildup of charge along the right side of the strip (and a corresponding deficiency of charge of that sign on the opposite side of the strip), which is the Hall effect, produces an electric field $\vec{\mathbf{E}}_{\mathrm{H}}$ across the strip, as shown in Fig. 32-18*b*. Equivalently, a potential difference $\Delta V_{\mathrm{H}} = E_{\mathrm{H}}w$, called the *Hall potential difference* (or Hall voltage), exists across the strip. We can measure ΔV_{H} by connecting the leads of a voltmeter to points x and y of Fig. 32-18. As we show below, the sign of ΔV_{H} gives the sign of the charge carriers, and the magnitude of ΔV_{H} gives their density (number per unit volume). If the charge carriers are electrons, for example, an excess of negative charges builds up on the right side of the strip, and point y is at a lower potential than point x. This may seem like an obvious conclusion in the case of metals; however, keep in mind that Hall's work was done nearly 20 years before Thomson's discovery of the electron, and the nature of electrical conduction in metals was not at all obvious at that time.

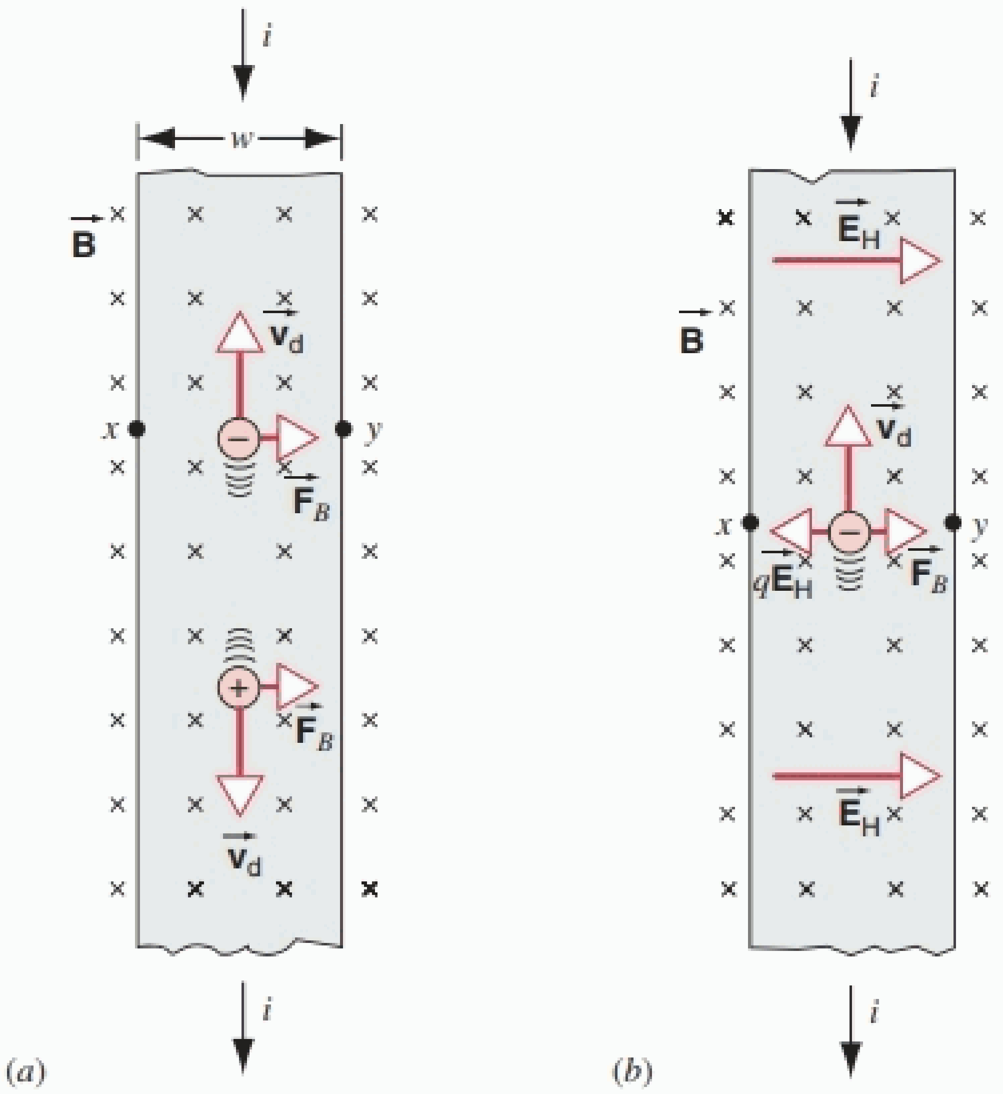


FIGURE 32-18. A strip of copper immersed in a magnetic field $\vec{\mathbf{B}}$ carries a current i. (*a*) The situation just after the magnetic field has been turned on, and (*b*) the situation at equilibrium, which quickly follows. Negative charges pile up on the right side of the strip, leaving uncompensated positive charges on the left. Point x is at a higher potential than point y.

Let us assume that conduction in the material is due to charge carriers of a particular sign (positive or negative) moving with drift velocity $\vec{\mathbf{v}}_{\mathrm{d}}$. As the charge carriers drift, they are deflected to the right in Fig. 32-18 by the magnetic force. As the charges collect on the right side, they set up an electric field that acts inside the conductor to oppose the sideways motion of additional charge carriers. Equilibrium is quickly reached, and the Hall voltage reaches its maximum; the sideways magnetic force ($\vec{\mathbf{F}}_B = q\vec{\mathbf{v}}_{\mathrm{d}} \times \vec{\mathbf{B}}$) is then balanced by the sideways electric force ($q\vec{\mathbf{E}}_{\mathrm{H}}$). In vector terms, the Lorentz force on the charge carriers under these circumstances is zero:

$$q\vec{\mathbf{E}}_{\mathrm{H}} + q\vec{\mathbf{v}}_{\mathrm{d}} \times \vec{\mathbf{B}} = 0, \tag{32-18}$$

or

$$\vec{\mathbf{E}}_{\mathrm{H}} = -\vec{\mathbf{v}}_{\mathrm{d}} \times \vec{\mathbf{B}}. \tag{32-19}$$

Since $\vec{\mathbf{v}}_{\mathrm{d}}$ and $\vec{\mathbf{B}}$ are at right angles, we can write Eq. 32-19 in terms of magnitudes as

$$E_{\mathrm{H}} = v_{\mathrm{d}}B. \tag{32-20}$$

From Eq. 29-6 we can write the drift speed as $v_{\mathrm{d}} = j/ne$, where j is the current density in the strip and n is the density of charge carriers. The current density j is the current i per unit cross-sectional area A of the strip. If t is the thickness of the strip, then its cross-sectional area A can be written as wt. Substituting $\Delta V_{\mathrm{H}}/w$ for the electric field E_{H}, we obtain

$$\frac{\Delta V_{\mathrm{H}}}{w} = v_{\mathrm{d}}B = \frac{j}{ne}B = \frac{i}{wtne}B$$

or, solving for the density of charge carriers,

$$n = \frac{iB}{et\,\Delta V_{\mathrm{H}}}. \tag{32-21}$$

From a measurement of the magnitude of the Hall potential difference ΔV_{H}, we can find the number density of the charge carriers. Table 32-2 shows a summary of Hall effect data for several metals and semiconductors. For some

TABLE 32-2 **Hall Effect Results for Selected Materials**

Material	n (10^{28}/m^3)	*Sign of* ΔV_{H}	*Number per atom*[a]
Na	2.5	−	0.99
K	1.5	−	1.1
Cu	11	−	1.3
Ag	7.4	−	1.3
Al	21	−	3.5
Sb	0.31	−	0.09
Be	2.6	+	2.2
Zn	19	+	2.9
Si (pure)	1.5×10^{-12}	−	3×10^{-13}
Si (typical n-type)	10^{-7}	−	2×10^{-8}

[a] The number of charge carriers per atom of the material as determined from the number per unit volume and the density and molar mass of the material.

monovalent metals (Na, K, Cu, Ag) the Hall effect indicates that each atom contributes approximately one free electron to the conduction. For other metals, the number of electrons can be greater than one per atom (Al) or less than one per atom (Sb). For some metals (Be, Zn), the Hall potential difference shows that the charge carriers have a *positive* sign. In this case the conduction is dominated by *holes,* unoccupied energy levels in the valence band (see Chapter 49). The holes correspond to the absence of an electron and thus behave like positive charge carriers moving through the material. For some materials, semiconductors in particular, there may be substantial contributions from both electrons and holes, and the simple interpretation of the Hall effect in terms of free conduction by one type of charge carrier is not sufficient. In this case we must use more detailed calculations based on quantum theory.

SAMPLE PROBLEM 32-3. A strip of copper 150 μm thick is placed in a magnetic field $B = 0.65$ T perpendicular to the plane of the strip, and a current $i = 23$ A is set up in the strip. What Hall potential difference ΔV_H would appear across the width of the strip if there were one charge carrier per atom?

Solution In Sample Problem 29-3, we calculated the number of charge carriers per unit volume for copper, assuming that each atom contributes one electron, and we found $n = 8.49 \times 10^{28}$ electrons/m^3. From Eq. 32-21 then,

$$\begin{aligned}\Delta V_H &= \frac{iB}{net}\\ &= \frac{(23\text{ A})(0.65\text{ T})}{(8.49 \times 10^{28}\text{ m}^{-3})(1.60 \times 10^{-19}\text{ C})(150 \times 10^{-6}\text{ m})}\\ &= 7.3 \times 10^{-6}\text{ V} = 7.3\ \mu\text{V}.\end{aligned}$$

This potential difference, though small, is readily measurable.

The Quantized Hall Effect* (Optional)

Let us rewrite Eq. 32-21 as

$$\frac{\Delta V_H}{i} = \frac{1}{etn}B. \tag{32-22}$$

The quantity on the left has the dimension of resistance (voltage divided by current), although it is not a resistance in the conventional sense. It is commonly called the *Hall resistance.* We can determine the Hall resistance by measuring the Hall voltage ΔV_H in a material carrying a current i.

Equation 32-22 shows that the Hall resistance is expected to increase linearly with the magnetic field B for a particular sample of material (in which n and t are constants). A plot of the Hall resistance against B should be a straight line.

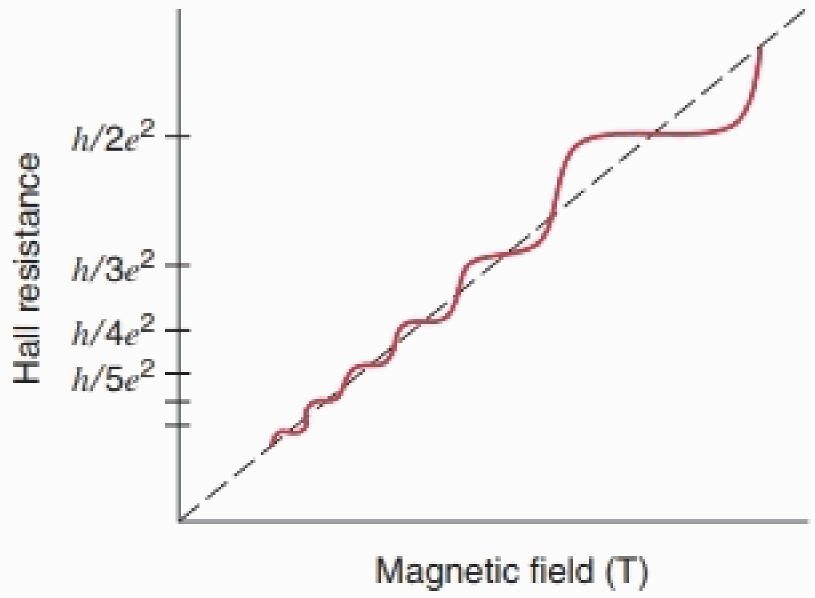


FIGURE 32-19. The quantized Hall effect. The dashed line shows the expected classical behavior. The steps show the quantum behavior.

In experiments done in 1980, German physicist Klaus von Klitzing discovered that, at high magnetic fields and low temperatures (about 1 K), the Hall resistance did not increase linearly with the field; instead, the plot showed a series of "stair steps," as shown in Fig. 32-19. This effect has become known as the *quantized Hall effect,* and von Klitzing was awarded the 1985 Nobel Prize in physics for his discovery.

The explanation for this effect involves the circular paths in which electrons are forced to move by the field. Quantum mechanics prevents the electron orbits of neighboring atoms from overlapping. As the field increases, the orbital radius decreases, permitting more orbits to bunch together on one side of the material. Because the orbital motion of electrons is quantized (only certain orbits are allowed), the changes in orbital motion occur suddenly, corresponding to the steps in Fig. 32-19. A natural unit of resistance corresponding to orbital motion is h/e^2, where h is the Planck constant, and the steps in Fig. 32-19 occur at Hall resistances of $h/2e^2$, $h/3e^2$, $h/4e^2$, and so on.

The quantized Hall resistance h/e^2, now known as the von Klitzing constant, has the value 25812.807572 Ω and has been determined to a precision of better than 1 part in 10^8, so the quantized Hall effect has provided a new standard for resistance. This standard, which can be duplicated exactly in laboratories around the world, became the new representation for the ohm in 1990. ■

32-5 THE MAGNETIC FORCE ON A CURRENT-CARRYING WIRE

A current is a collection of moving charges. Because a magnetic field exerts a sideways force on a moving charge, it should also exert a sideways force on a wire carrying a current. That is, a sideways force is exerted on the conduction electrons in the wire, but since the electrons cannot escape sideways, the force must be transmitted to the wire itself. Figure 32-20 shows a wire that passes through a region in which a magnetic field $\vec{\mathbf{B}}$ exists. When the wire carries

---

*See "The Quantized Hall Effect," by Bertrand I. Halperin, *Scientific American,* April 1986, p. 52.

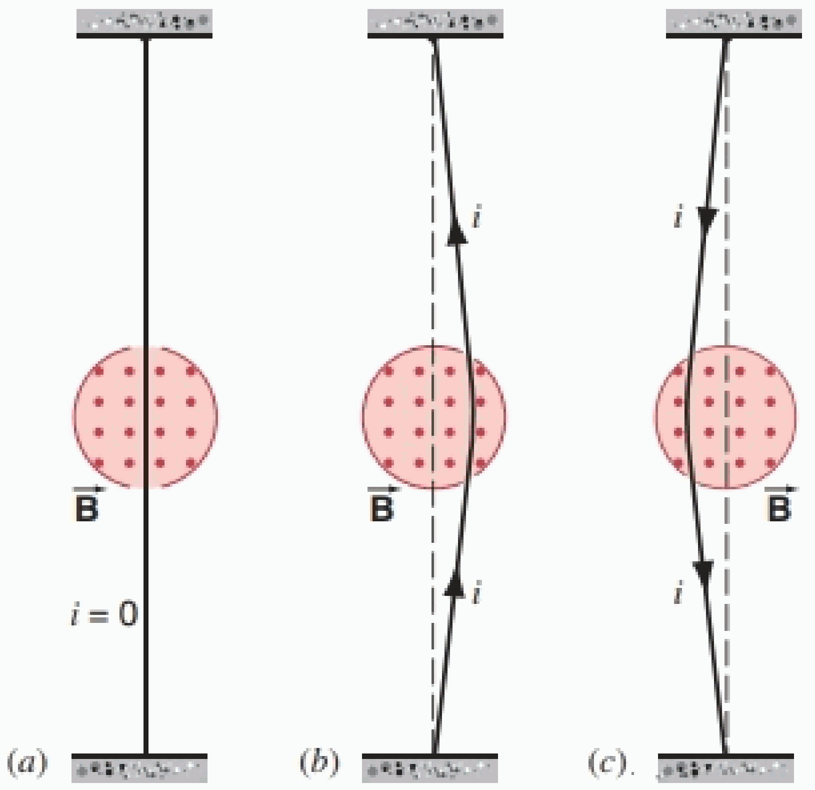


FIGURE 32-20. A flexible wire passes between the poles of a magnet. (*a*) There is no current in the wire. (*b*) A current is established in the wire. (*c*) The current is reversed.

no current (Fig. 32-20*a*), it experiences no deflection. When a current is carried by the wire, it deflects (Fig. 32-20*b*); when the current is reversed (Fig. 32-20*c*), the deflection reverses. The deflection also reverses when the field $\vec{\mathbf{B}}$ is reversed.

To understand this effect, we consider the individual charges flowing in a wire (Fig. 32-21). We use the free-electron model (Section 29-3) for current in a wire, assuming the electrons to move with a constant velocity, the drift velocity $\vec{\mathbf{v}}_d$. The actual direction of motion of the electrons is of course opposite to the direction we take for the current i in the wire.

The wire passes through a region in which a uniform field $\vec{\mathbf{B}}$ exists. The sideways force on each electron (of charge $q = -e$) due to the magnetic field is $-e\vec{\mathbf{v}}_d \times \vec{\mathbf{B}}$. Let us consider the total sideways force on a segment of the wire of length L. The same force (magnitude and direction) acts on each electron in the segment, and the total force $\vec{\mathbf{F}}_B$ on the segment is therefore equal to the number N of electrons times the force on each electron:

$$\vec{\mathbf{F}}_B = -Ne\vec{\mathbf{v}}_d \times \vec{\mathbf{B}}. \tag{32-23}$$

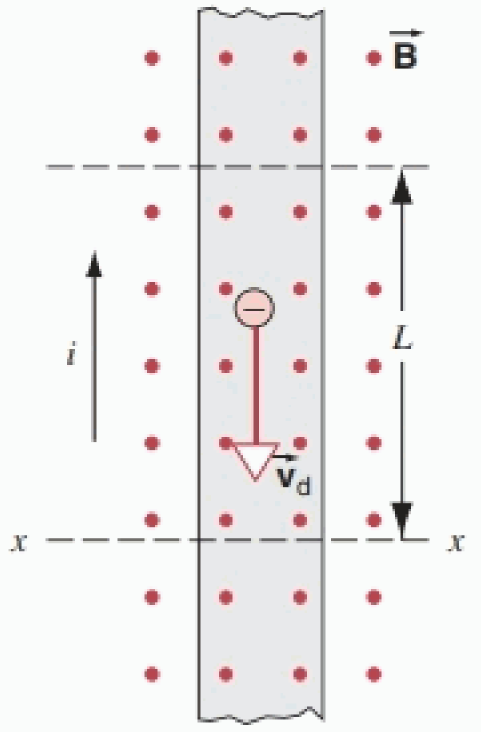


FIGURE 32-21. A close-up view of a length L of the wire of Fig. 32-20*b*. The current direction is upward, which means that electrons drift downward. A magnetic field emerges from the plane of the figure, so that the wire is deflected to the right.

How many electrons are contained in that segment of wire? If n is the number density (number per unit volume) of electrons, then the total number N of electrons in the segment is nAL, where A is the cross-sectional area of the wire. Substituting into Eq. 32-23, we obtain

$$\vec{\mathbf{F}}_B = -nALe\vec{\mathbf{v}}_d \times \vec{\mathbf{B}}. \tag{32-24}$$

Equation 29-6 permits us to write Eq. 32-24 in terms of the current i. To preserve the vector relationship of Eq. 32-24, we define the vector $\vec{\mathbf{L}}$ to be equal in magnitude to the length of the segment and to point in the direction of the current (opposite to the direction of electron flow). The vectors $\vec{\mathbf{v}}_d$ and $\vec{\mathbf{L}}$ have opposite directions, and we can write the scalar relationship $nALev_d = iL$ using vectors as

$$-nALe\vec{\mathbf{v}}_d = i\vec{\mathbf{L}}. \tag{32-25}$$

Substituting Eq. 32-25 into Eq. 32-24, we obtain an expression for the force on the segment:

$$\vec{\mathbf{F}}_B = i\vec{\mathbf{L}} \times \vec{\mathbf{B}}. \tag{32-26}$$

Equation 32-26 is similar to Eq. 32-3 ($\vec{\mathbf{F}}_B = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}$), in that either can be taken to be the defining equation for the magnetic field. Figure 32-22 shows the vector relationship between $\vec{\mathbf{F}}$, $\vec{\mathbf{L}}$, and $\vec{\mathbf{B}}$; compare with Fig. 32-5 to see the similarities between Eqs. 32-26 and 32-3.

If the field is uniform over the length of the wire segment and the direction of the current makes an angle ϕ with the field, then the magnitude of the force is (compare Eq. 32-2)

$$F_B = iLB \sin \phi. \tag{32-27}$$

If $\vec{\mathbf{L}}$ is parallel to $\vec{\mathbf{B}}$, then the force is zero. If the segment is perpendicular to the direction of the field, the magnitude of the force is

$$F_B = iLB. \tag{32-28}$$

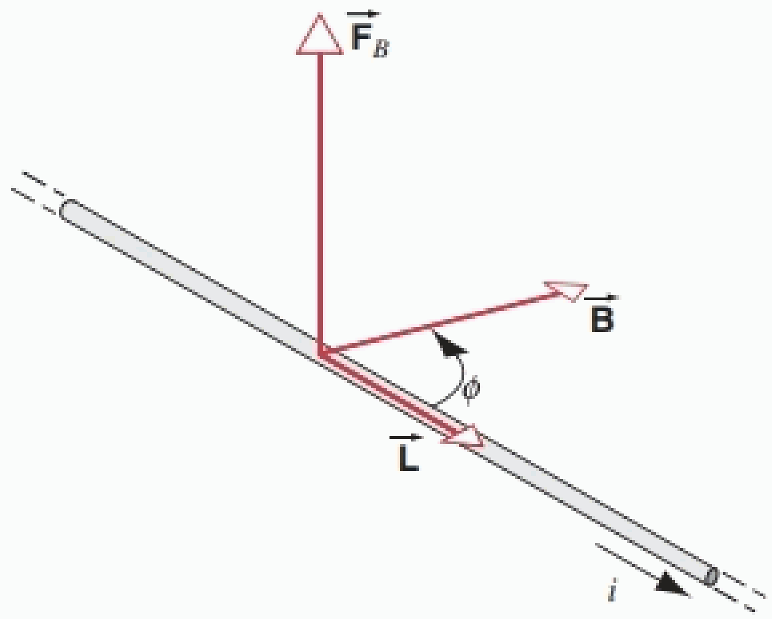


FIGURE 32-22. The magnetic force acting on a directed wire segment $\vec{\mathbf{L}}$ that makes an angle ϕ with a magnetic field. Compare carefully with Fig. 32-5.

If the wire is not straight or the field is not uniform, we can imagine the wire to be broken into small segments of length dL; we make the segments small enough that they are approximately straight and the field is approximately uniform. The force on each segment can then be written

$$d\vec{\mathbf{F}}_B = i\, d\vec{\mathbf{L}} \times \vec{\mathbf{B}}. \tag{32-29}$$

We can find the total force on the segment of length L by doing a suitable integration over the length.

SAMPLE PROBLEM 32-4. A straight, horizontal segment of copper wire carries a current $i = 28$ A. What are the magnitude and direction of the magnetic field needed to "float" the wire—that is, to balance its weight? Its linear mass density is 46.6 g/m.

Solution Figure 32-23 shows the arrangement. For a length L of wire we have (see Eq. 32-28)

$$mg = iLB,$$

or

$$B = \frac{(m/L)g}{i} = \frac{(46.6 \times 10^{-3}\ \text{kg/m})(9.8\ \text{m/s}^2)}{28\ \text{A}}$$
$$= 1.6 \times 10^{-2}\ \text{T} = 16\ \text{mT}.$$

This is about 400 times the strength of the Earth's magnetic field.

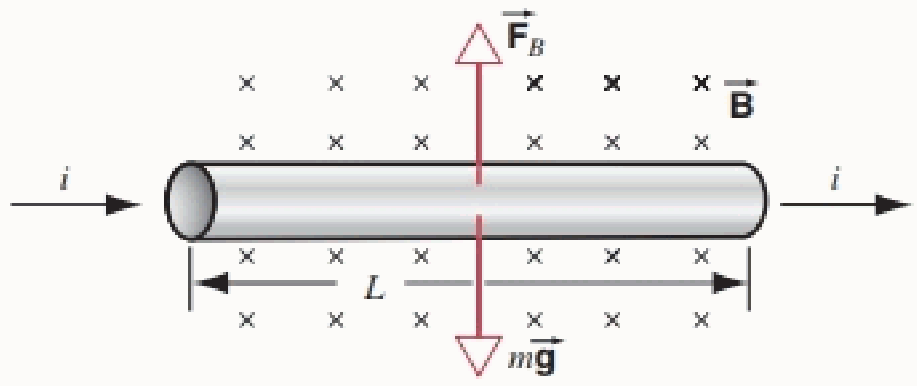


FIGURE 32-23. Sample Problem 32-4. A wire can be made to "float" in a magnetic field, with the upward magnetic force $\vec{\mathbf{F}}_B$ balancing the downward pull of gravity. The magnetic field is into the plane of the page.

SAMPLE PROBLEM 32-5. Figure 32-24 shows a wire segment, placed in a uniform magnetic field $\vec{\mathbf{B}}$ that points out of the plane of the figure. If the segment carries a current i, what resultant magnetic force acts on it?

Solution Let us consider the wire in the three sections—the two straight parts (secctions 1 and 3) and the curved part (section 2). According to Eq. 32-28, the magnetic force that acts on each straight section has the magnitude

$$F_1 = F_3 = iLB$$

and points down, as shown by the arrows in the figure. The force dF_B that acts on a segment of the arc of length $dL = R\, d\theta$ has magnitude

$$dF_B = iB\, ds = iB(R\, d\theta)$$

and direction radially toward O, the center of the arc. Note that only the downward component ($dF_B \sin\theta$) of this force element is effective. The horizontal component ($dF_B \cos\theta$) is cancelled by an oppositely directed horizontal component due to a symmetrically located segment on the opposite side of the arc.

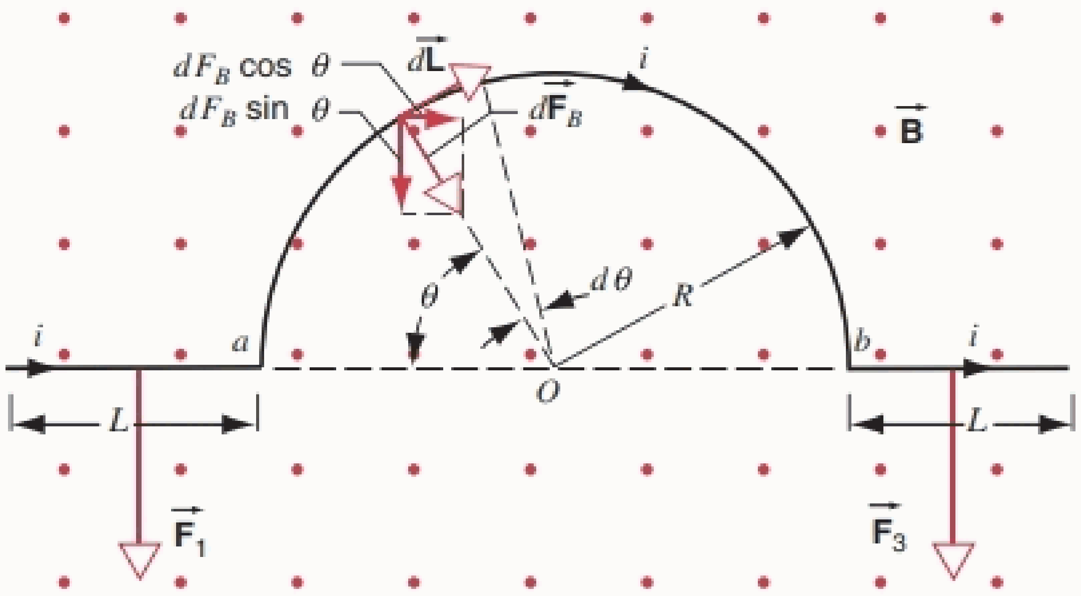


FIGURE 32-24. Sample Problem 32-5. A wire segment carrying a current i is immersed in a magnetic field. The resultant force on the wire is directed downward.

The total force on the central arc points down and is given by

$$F_2 = \int_0^\pi dF_B \sin\theta = \int_0^\pi (iBR\, d\theta) \sin\theta$$
$$= iBR \int_0^\pi \sin\theta\, d\theta = 2iBR.$$

The resultant force on the entire wire is then

$$F = F_1 + F_2 + F_3 = iLB + 2iBR + iLB$$
$$= iB(2L + 2R).$$

The same force would also act on a wire similar to that of Fig. 32-24, with the central semicircular segment replaced by a segment of any shape (including a straight line) connecting points a and b. Can you convince yourself that this is so?

32-6 THE TORQUE ON A CURRENT LOOP

In an electric motor, a loop of wire carrying a current is placed in a magnetic field. A simplified version is shown in Fig. 32-25 for a rectangular loop in a uniform field. The loop is free to pivot about a vertical axis. When the loop is oriented so that the field is in the plane of the loop, the magnetic forces on the short ends are zero according to Eq. 32-26, because $\vec{\mathbf{B}}$ and $\vec{\mathbf{L}}$ are parallel. On the long ends of the rectangular loop, the forces are equal but point in opposite directions, so the net force on the loop is zero. However, there is a net torque that tends to rotate the loop about its axis in a clockwise direction when viewed from above. This simple model shows how the combination of an electric current and a magnetic field can produce the rotary motion of the electric motor. The same principle is responsible for the action of analog voltmeters and ammeters.

Figure 32-26 shows a rectangular loop of length a and width b carrying a current i. The plane of the loop makes an angle θ with the x axis. For simplicity, only the loop itself is

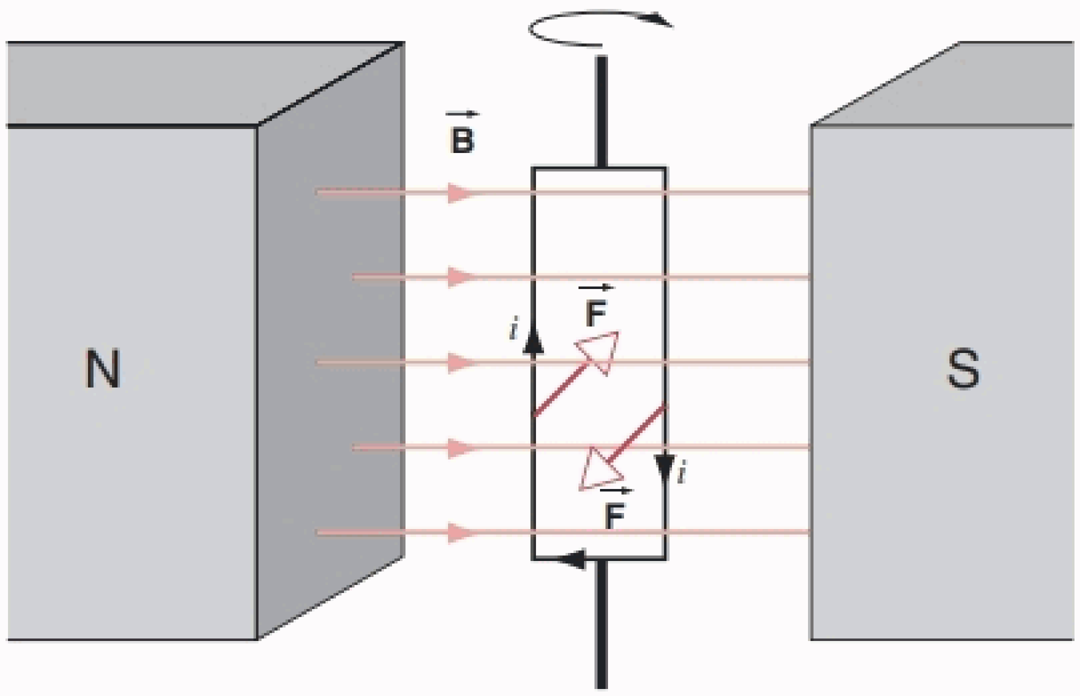


FIGURE 32-25. A simplified diagram of an electric motor. The loop carries an electric current. The magnetic forces on the long sides of the loop produce a torque that tends to rotate the loop clockwise as viewed from above.

shown; wires needed to bring current into and out of the loop are not shown. The magnetic field is taken to be uniform and in the y direction, and the z axis lies in the plane of the loop. Our goal is to find the net force and the net torque on the loop by calculating the force on each side of the loop.

In this orientation, sides 1 and 3 are perpendicular to the field. That is, if we define the vector $\vec{\mathbf{L}}$ as being in the direction of the current, then $\vec{\mathbf{L}}$ is perpendicular to $\vec{\mathbf{B}}$. For those sides, we can use Eq. 32-28 for the magnitudes of the forces:

$$F_1 = F_3 = iaB, \qquad (32\text{-}30)$$

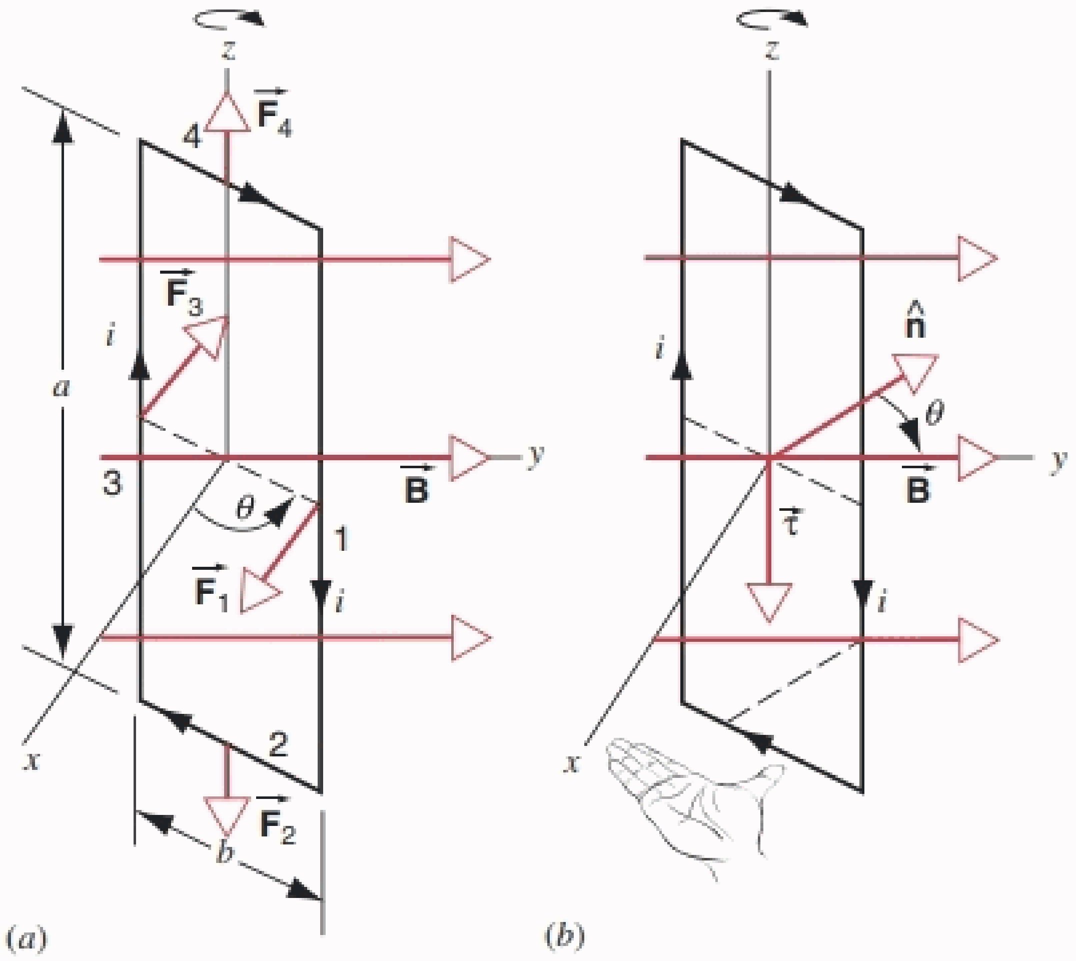


FIGURE 32-26. A rectangular loop of wire in a uniform magnetic field. (*a*) The forces on the four sides are shown. (*b*) The torque tends to rotate the loop so that the unit vector $\hat{\mathbf{n}}$, determined from the right-hand rule and perpendicular to the plane of the loop, rotates into alignment with $\vec{\mathbf{B}}$.

because sides 1 and 3 each have length a. The forces are parallel to the x axis of Fig. 32-26, with $\vec{\mathbf{F}}_1$ in the positive x direction and $\vec{\mathbf{F}}_3$ in the negative x direction.

The angle between side 2 of the wire and $\vec{\mathbf{B}}$ is $\theta + 90°$. Using Eq. 32-27, we find the force on this segment to be

$$F_2 = ibB \sin(\theta + 90°) = ibB \cos\theta \qquad (32\text{-}31)$$

in the negative z direction. Similarly, the force on side 4 is

$$F_4 = ibB \sin(90° - \theta) = ibB \cos\theta \qquad (32\text{-}32)$$

in the positive z direction.

To find the total force on the loop, we add the forces on the four sides, being careful to take into account both their magnitudes and directions. Because F_2 and F_4 are equal in magnitude and opposite in direction, they sum to zero; the same is true for F_1 and F_3. The net force on the loop is zero, so its center of mass does not accelerate under the action of the magnetic force. This conclusion follows because the field is uniform; if the field were nonuniform, the field at opposite pairs of sides 1 and 3 or 2 and 4 might have different magnitudes, and the forces on those sides might not be equal in magnitude.

Even though the net force is zero, the net torque is nonzero. Forces F_2 and F_4 both lie along the z axis and so have the same line of action; they do not contribute to the net torque. However, F_1 and F_3 do not have the same line of action; as you can see from Fig. 32-26, they tend to rotate the loop clockwise about the z axis as viewed from above. Relative to the z axis, the forces F_1 and F_3 each have moment arms of $(b/2)\sin\theta$, and so the magnitude of the total torque is

$$\tau = 2(iaB)(b/2)\sin\theta, \qquad (32\text{-}33)$$

where the factor 2 enters because both forces contribute equally to the torque.

The torque has its maximum magnitude when the loop is oriented so that the magnetic field lies in the plane of the loop ($\theta = 90°$). The torque is zero when the magnetic field is perpendicular to the plane of the loop ($\theta = 0$).

If the loop were constructed as a coil of N turns of wire (such as might be found in a motor or an ammeter), Eq. 32-33 gives the torque on each turn, and the total torque on the coil would be

$$\tau = NiAB\sin\theta, \qquad (32\text{-}34)$$

where we have substituted A, the area of the loop, for the product ab. Equation 32-34 can be shown to hold in general for all plane loops of area A whether or not they are rectangular.

Figure 32-26*b* gives another way to interpret the torque on the current loop. Using the right-hand rule, we define a unit vector $\hat{\mathbf{n}}$ perpendicular to the plane of the loop. The direction of $\hat{\mathbf{n}}$ can be determined by holding your right hand so the fingers follow the direction of the current; your thumb indicates the direction of $\hat{\mathbf{n}}$. The torque tries to rotate the loop so that $\hat{\mathbf{n}}$ is brought into alignment with $\vec{\mathbf{B}}$. The torque, which is in the negative z direction in Fig. 32-26*b*, is in the direction determined by the cross product $\hat{\mathbf{n}} \times \vec{\mathbf{B}}$. With $|\hat{\mathbf{n}} \times \vec{\mathbf{B}}| = B \sin\theta$, we can write Eq. 32-34 in vector form as

$$\vec{\boldsymbol{\tau}} = NiA\,\hat{\mathbf{n}} \times \vec{\mathbf{B}}. \qquad (32\text{-}35)$$

SAMPLE PROBLEM 32-6. Analog voltmeters and ammeters, in which the reading is displayed by the deflection of a pointer over a scale, work by measuring the torque exerted by a magnetic field on a current loop. Figure 32-27 shows the rudiments of a *galvanometer*, on which both analog ammeters and analog voltmeters are based. The coil is 2.1 cm high and 1.2 cm wide; it has 250 turns and is mounted so that it can rotate about its axis in a uniform radial magnetic field with $B = 0.23$ T. A spring provides a countertorque that balances the magnetic torque, resulting in a steady angular deflection ϕ corresponding to a given steady current i in the coil. If a current of 100 μA produces an angular deflection of 28° $(= 0.49 \text{ rad})$, what must be the torsional constant κ of the spring?

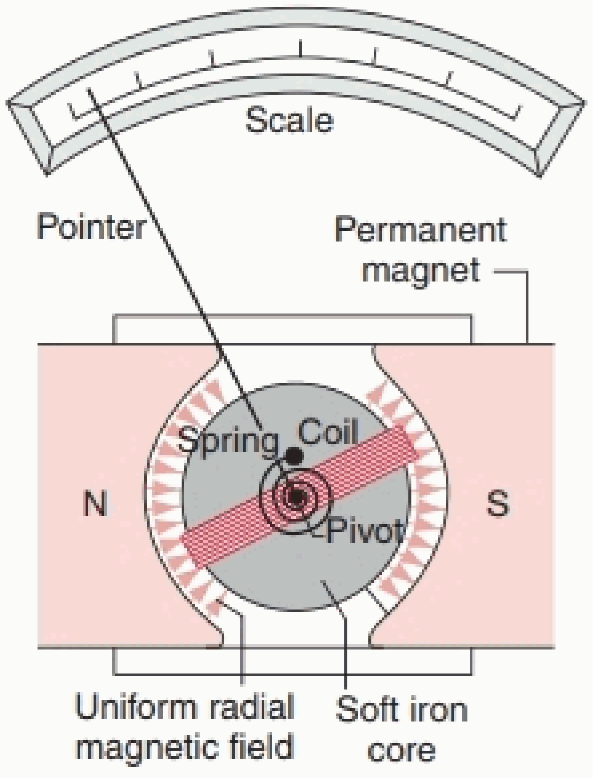


FIGURE 32-27. Sample Problem 32-6. The rudiments of a galvanometer. Depending on the external circuit, this device can act as either a voltmeter or an ammeter.

Solution Setting the magnetic torque (Eq. 32-34) equal to the restoring torque $\kappa\phi$ of the spring yields

$$\tau = NiAB \sin\theta = \kappa\phi,$$

in which ϕ is the angular deflection of the pointer and A $(= 2.52 \times 10^{-4} \text{ m}^2)$ is the area of the coil. Note that the normal to the plane of the coil (that is, the pointer) is always at right angles to the (radial) magnetic field so that $\theta = 90°$ for all pointer positions.

Solving for κ, we find

$$\kappa = \frac{NiAB \sin\theta}{\phi}$$

$$= \frac{(250)(100 \times 10^{-6} \text{ A})(2.52 \times 10^{-4} \text{ m}^2)(0.23 \text{ T})(\sin 90°)}{0.49 \text{ rad}}$$

$$= 3.0 \times 10^{-6} \text{ N} \cdot \text{m/rad}.$$

Modern ammeters and voltmeters are of the digital, direct-reading type and operate in a way that does not involve a moving coil.

MULTIPLE CHOICE

32-1 Magnetic Interactions and Magnetic Poles

32-2 The Magnetic Force on a Moving Charge

1. Of the three vectors in the equation $\vec{\mathbf{F}}_B = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}$, which pair(s) are always at right angles? *(There may be more than one correct answer.)*

(A) $\vec{\mathbf{F}}_B$ and $\vec{\mathbf{v}}$ (B) $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$ (C) $\vec{\mathbf{B}}$ and $\vec{\mathbf{F}}_B$
(D) None (E) All three must be at right angles.

2. A negative charge q_1 is moving with a constant velocity $\vec{\mathbf{v}}$ in a region where both a uniform electric field $\vec{\mathbf{E}}$ and uniform magnetic field $\vec{\mathbf{B}}$ exist.

(*a*) Of the three vectors $\vec{\mathbf{v}}$, $\vec{\mathbf{E}}$, and $\vec{\mathbf{B}}$, which pair(s) must be perpendicular? *(There may be more than one correct answer.)*

(A) $\vec{\mathbf{E}}$ and $\vec{\mathbf{v}}$ (B) $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$ (C) $\vec{\mathbf{B}}$ and $\vec{\mathbf{E}}$
(D) None (E) All three must be perpendicular.

(*b*) The negative charge is replaced with another charge q_2 moving initially with the same velocity. Under what conditions will the second charge also move with constant velocity?

(A) q_2 must be identical to q_1.
(B) q_2 must be negative, but can have any magnitude.
(C) q_2 can be positive, but must have the same magnitude as q_1.
(D) q_2 can be any charge.

3. An electron is released from rest in a region of space where a uniform electric field $\vec{\mathbf{E}}$ exists pointing *up* the page and a uniform magnetic field $\vec{\mathbf{B}}$ is pointing *out* of the page. Which path in Fig. 32-28 best represents the motion of the electron after it is released?

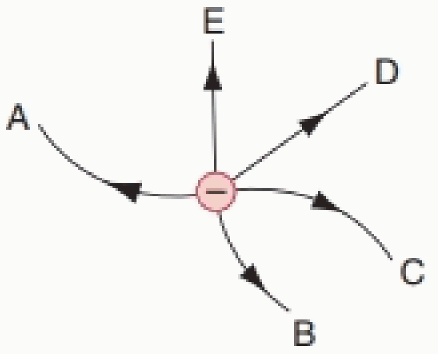


FIGURE 32-28. Multiple-choice question 3.

4. Which of the following properties of a proton can change while it moves freely in a uniform electric field $\vec{\mathbf{E}}$? *(There may be more than one correct answer.)*

(A) Mass (B) Speed (C) Velocity
(D) Momentum (E) Kinetic energy

5. Which of the following properties of a proton can change while it moves freely in a uniform magnetic field $\vec{\mathbf{B}}$? *(There may be more than one correct answer.)*
(A) Mass (B) Speed (C) Velocity
(D) Momentum (E) Kinetic energy

6. Which of the following properties of a proton can change while it moves freely in a *non*uniform magnetic field $\vec{\mathbf{B}}$? *(There may be more than one correct answer.)*
(A) Mass (B) Speed (C) Velocity
(D) Momentum (E) Kinetic energy

7. Can a static magnetic field do positive work on a charged particle?
(A) Yes
(B) Yes, but only if the particle has a positive charge
(C) Yes, but only if the particle has an initial velocity
(D) No

8. A region of space has a uniform electric field $\vec{\mathbf{E}}$ directed down and a uniform magnetic field $\vec{\mathbf{B}}$ directed east. Gravity is negligible. An electron is moving with a constant velocity $\vec{\mathbf{v}}_1$ through the two fields.

(*a*) In which direction could the electron be moving? *(There may be more than one correct answer.)*
(A) North (B) South (C) Up (D) Down

(*b*) A second electron originally follows the *direction* of the first, but is moving at a slightly slower speed $v_2 < v_1$. What is the direction of the net force on the second electron?
(A) North (B) South (C) Up (D) Down

32-3 Circulating Charges

9. An electron with a speed $v_0 \ll c$ moves in a circle of radius r_0 in a uniform magnetic field. The time required for one revolution of the electron is T_0. The speed of the electron is now doubled to $2v_0$.

(*a*) The radius of the circle will change to
(A) $4r_0$. (B) $2r_0$. (C) r_0. (D) $r_0/2$.

(*b*) The time required for one revolution of the electron will change to
(A) $4T_0$. (B) $2T_0$. (C) T_0. (D) $T_0/2$.

10. Consider the motion of the charge in Fig. 32-17. Describe the magnetic field.
(A) The field is strongest near the center.
(B) The field is weakest near the center.
(C) There is not enough information to solve the problem.

32-4 The Hall Effect

11. The magnetic field in Fig. 32-29 points into the page. A small metal plane moves down the page.

(*a*) According to the pilot, which wing becomes negatively charged while the plane is moving?
(A) The left wing (B) The right wing
(C) Neither wing becomes charged.
(D) The answer depends on the sign of the charge carriers in the plane.

(*b*) The plane is now turned around and "flies" to the top of the page. Which wing now becomes negatively charged while the plane moves up the page?
(A) The left wing (B) The right wing
(C) Neither wing becomes charged.
(D) The answer depends on the sign of the charge carriers in the plane.

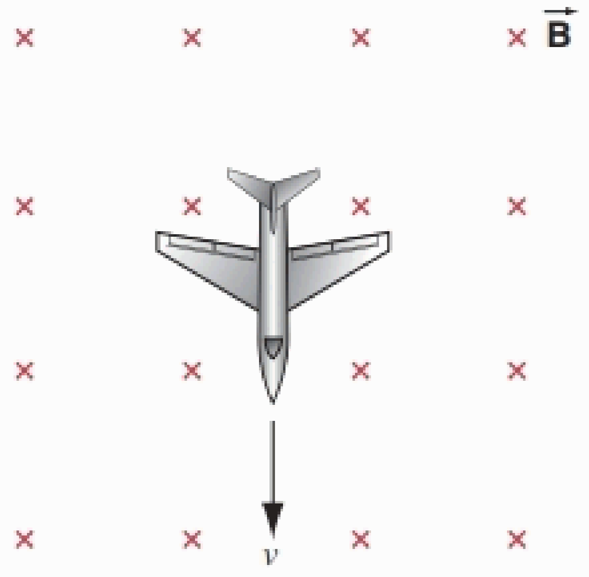


FIGURE 32-29. Multiple-choice question 11.

32-5 The Magnetic Force on a Current-Carrying Wire

12. Figure 32-30 shows several wire segments that carry equal currents from *a* to *b*. The wires are in a uniform magnetic field $\vec{\mathbf{B}}$ directed into the page. Which wire segment experiences the largest net force?
(A) 1 (B) 2 (C) 3
(D) All experience the same net force.
(E) The question cannot be answered without additional information.

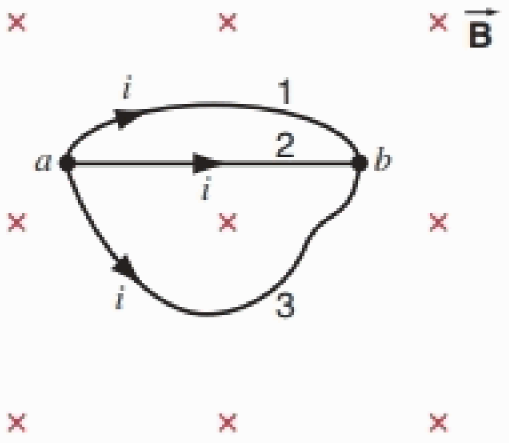


FIGURE 32-30. Multiple-choice questions 12 and 13.

13. Repeat Multiple-choice question 12, except assume that the uniform magnetic field could be pointing in *any* direction. Which wire segment experiences the largest net force?
(A) 1 (B) 2 (C) 3
(D) All experience the same net force.
(E) The answer depends on the direction of the magnetic field.

32-6 The Torque on a Current Loop

14. Does Eq. 32-35 hold true for single-turn wire loops in shapes other than rectangles?
(A) It is a reasonable approximation for shapes close to rectangular.
(B) It is a reasonable approximation for any shape that lies in a plane.
(C) It is true for shapes of sufficient symmetry, such as equilateral triangles or circles.
(D) It is true for all shapes that lie in a plane.

15. The wire loop in Fig. 32-31 carries a clockwise current. There is a uniform magnetic field $\vec{\mathbf{B}}$ directed to the right.

(*a*) What is the direction of the net force on the current loop?
(A) Into the page (B) Out of the page
(C) Up the page (D) Down the page
(E) The net force is zero.

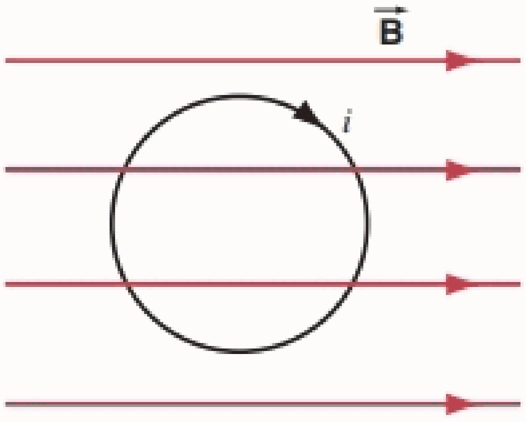


FIGURE 32-31. Multiple-choice question 15.

(*b*) What is the direction of the torque on the current loop?
(A) Into the page (B) Out of the page
(C) Up the page (D) Down the page
(E) The net force is zero.

16. Repeat Multiple-choice question 15, but now assume that the field is not uniform; it is stronger at the top of the page than at the bottom.

(*a*) What is the direction of the net force on the current loop?
(A) Into the page (B) Out of the page
(C) Up the page (D) Down the page
(E) The net force is zero.

(*b*) What is the direction of the torque on the current loop?
(A) Into the page (B) Out of the page
(C) Up the page (D) Down the page
(E) The net force is zero.

QUESTIONS

1. Why do we not simply define the direction of the magnetic field $\vec{\mathbf{B}}$ to be the direction of the magnetic force that acts on a moving charge?

2. Imagine that you are sitting in a room with your back to one wall and that an electron beam, traveling horizontally from the back wall to the front wall, is deflected to your right. What is the direction of the uniform magnetic field that exists in the room?

3. How could we rule out that the forces between two magnets are electrostatic forces?

4. If an electron is not deflected in passing through a certain region of space, can we be sure that there is no magnetic field in that region?

5. If a moving electron is deflected sideways in passing through a certain region of space, can we be sure that a magnetic field exists in that region?

6. A beam of electrons can be deflected either by an electric field or by a magnetic field. Is one method better than the other? In any sense easier?

7. Electric fields can be represented by maps of equipotential surfaces. Can the same be done for magnetic fields? Explain.

8. Is a magnetic force conservative or nonconservative? Justify your answer. Could we define a magnetic potential energy as we defined electric or gravitational potential energy?

9. A charged particle passes through a magnetic field and is deflected. This means that a force acted on it and changed its momentum. Where there is a force there must be a reaction force. On what object does it act?

10. In the Thomson experiment we neglected the deflections produced by the gravitational field and magnetic field of the Earth. What errors are thereby introduced?

11. Imagine the room in which you are seated to be filled with a uniform magnetic field pointing vertically downward. At the center of the room two electrons are suddenly projected horizontally with the same initial speed but in opposite directions. (*a*) Describe their motions. (*b*) Describe their motions if one particle is an electron and one a positron—that is, a positively charged electron. (The electrons will gradually slow down as they collide with molecules of the air in the room.)

12. Figure 32-32 shows the tracks of two electrons (e^-) and a positron (e^+) in a bubble chamber. A magnetic field fills the chamber, perpendicular to the plane of the figure. Why are the tracks spirals and not circles? What can you tell about the particles from their tracks? What is the direction of the magnetic field?

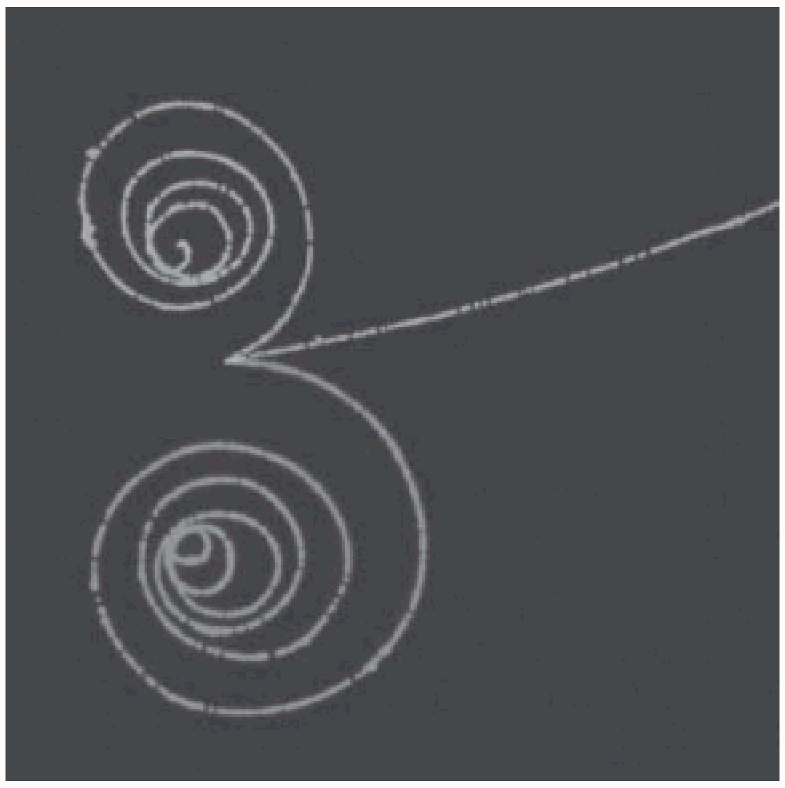


FIGURE 32-32. Question 12.

13. What are the primary functions of (*a*) the electric field and (*b*) the magnetic field in the cyclotron?

14. In a given magnetic field, would a proton or an electron, traveling at the same speed, have the greater frequency of revolution? Consider relativistic effects.

15. What central fact makes the operation of a conventional cyclotron possible? Ignore relativistic considerations.

16. A bare copper wire emerges from one wall of a room, crosses the room, and disappears into the opposite wall. You are told that there is a steady current in the wire. How can you find its direction? Describe as many ways as you can think of. You may use any reasonable piece of equipment, but you may not cut the wire.

17. Discuss the possibility of using the Hall effect to measure the strength B of a magnetic field.

18. (*a*) In measuring Hall potential differences, why must we be careful that points x and y in Fig. 32-18 are exactly opposite each other? (*b*) If one of the contacts is movable, what procedure might we follow in adjusting it to make sure that the two points are properly located?

19. In Section 32-5, we state that a magnetic field $\vec{B}$ exerts a sideways force on the conduction electrons in, say, a copper wire carrying a current i. We have tacitly assumed that this same force acts on the conductor itself. Are there some missing steps in this argument? If so, supply them.

20. A straight copper wire carrying a current i is immersed in a magnetic field $\vec{B}$, at right angles to it. We know that $\vec{B}$ exerts a sideways force on the free (or conduction) electrons. Does it do so on the bound electrons? After all, they are not at rest. Discuss.

21. Does Eq. 32-26 ($\vec{F}_B = i\vec{L} \times \vec{B}$) hold for a straight wire whose cross section varies irregularly along its length (i.e., a "lumpy" wire)?

22. A current in a magnetic field experiences a force. Therefore, it should be possible to pump conducting liquids by sending a current through the liquid (in an appropriate direction) and letting it pass through a magnetic field. Design such a pump. This principle is used to pump liquid sodium (a conductor, but highly corrosive) in some nuclear reactors, where it is used as a coolant. What advantages would such a pump have?

23. A uniform magnetic field fills a certain cubical region of space. Can an electron be fired into this cube from the outside in such a way that it will travel in a closed circular path inside the cube?

24. A conductor, even though it is carrying a current, has zero net charge. Why then does a magnetic field exert a force on it?

25. A rectangular current loop is in an arbitrary orientation in an external magnetic field. How much work is required to rotate the loop about an axis perpendicular to its plane?

26. Equation 32-35 shows that there is no torque on a current loop in an external magnetic field if the angle between the axis of the loop and the field is (*a*) 0° or (*b*) 180°. Discuss the nature of the equilibrium (that is, whether it is stable, neutral, or unstable) for these two positions.

27. Imagine that the room in which you are seated is filled with a uniform magnetic field pointing vertically upward. A circular loop of wire has its plane horizontal. For what direction of current in the loop, as viewed from above, will the loop be in stable equilibrium with respect to forces and torques of magnetic origin?

EXERCISES

32-1 Magnetic Interactions and Magnetic Poles

32-2 The Magnetic Force on a Moving Charge

1. Four particles follow the paths shown in Fig. 32-33 as they pass through the magnetic field there. What can one conclude about the charge of each particle?

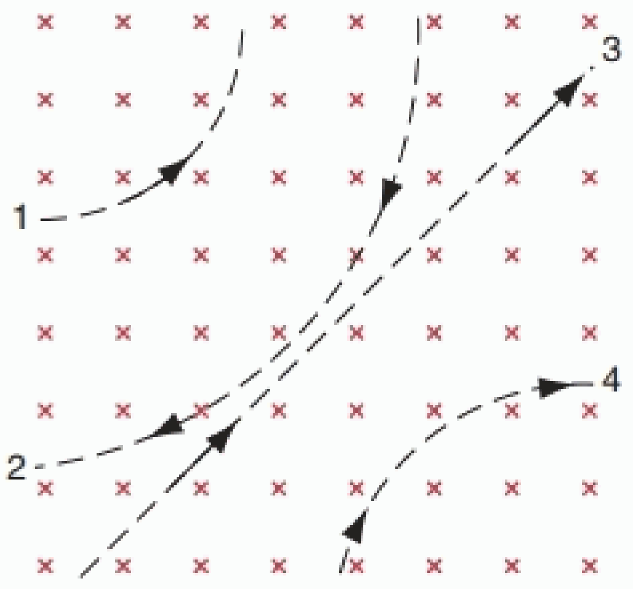


FIGURE 32-33. Exercise 1.

2. An electron in a TV camera tube is moving at 7.2×10^6 m/s in a magnetic field of strength 83 mT. (*a*) Without knowing the direction of the field, what could be the greatest and least magnitudes of the force the electron could feel due to the field? (*b*) At one point the acceleration of the electron is 4.9×10^{16} m/s^2. What is the angle between the electron's velocity and the magnetic field?

3. An electric field of 1.5 kV/m and a magnetic field of 0.44 T act on a moving electron to produce no force. (*a*) Calculate the minimum electron speed v. (*b*) Draw the vectors $\vec{E}$, $\vec{B}$, and $\vec{v}$.

4. A proton traveling at 23.0° with respect to a magnetic field of strength 2.63 mT experiences a magnetic force of 6.48×10^{-17} N. Calculate (*a*) the speed and (*b*) the kinetic energy in eV of the proton.

5. A cosmic ray proton strikes the Earth near the equator with a vertical velocity of 2.8×10^7 m/s. Assume that the horizontal component of the Earth's magnetic field at the equator is 30 μT. Calculate the ratio of the magnetic force on the proton to the gravitational force on it.

6. An electron is accelerated through a potential difference of 1.0 kV and directed into a region between two parallel plates separated by 20 mm with a potential difference of 100 V between them. If the electron enters moving perpendicular to the electric field between the plates, what magnetic field is necessary perpendicular to both the electron path and the electric field so that the electron travels in a straight line?

7. A uniform electric field $\vec{E}$ is perpendicular to a uniform magnetic field $\vec{B}$. A proton moving with velocity $\vec{v}_p$ perpendicular to both fields experiences no net force. An electron moving with velocity $\vec{v}_e$ also experiences no net force. Show that the ratio of the proton's kinetic energy to that of the electron is m_p/m_e.

8. An ion source is producing ions of ^{6}Li (mass = 6.01 u) each carrying a net charge of $+e$. The ions are accelerated by a potential difference of 10.8 kV and pass horizontally into a region in which there is a vertical magnetic field $B = 1.22$ T. Calculate the strength of the horizontal electric field to be set up over the same region that will allow the ^{6}Li ions to pass through undeflected.

32-3 Circulating Charges

9. (*a*) In a magnetic field with $B = 0.50$ T, for what path radius will an electron circulate at a speed of $0.10c$? (*b*) What will be its kinetic energy in eV? Ignore the small relativistic effects.

10. A 1.22-keV electron is circulating in a plane at right angles to a uniform magnetic field. The orbit radius is 24.7 cm. Calculate (*a*) the speed of the electron, (*b*) the magnetic field, (*c*) the frequency of revolution, and (*d*) the period of the motion.

11. An electron is accelerated from rest by a potential difference of 350 V. It then enters a uniform magnetic field of magnitude 200 mT, its velocity at right angles to this field. Calculate (*a*) the speed of the electron and (*b*) the radius of its path in the magnetic field.

12. S. A. Goudsmit devised a method for measuring accurately the masses of heavy ions by timing their period of revolution in a known magnetic field. A singly charged ion of iodine makes 7.00 rev in a field of 45.0 mT in 1.29 ms. Calculate its mass, in atomic mass units. Actually, the mass measurements are carried out to much greater accuracy than these approximate data suggest.

13. An alpha particle ($q = +2e$, $m = 4.0$ u) travels in a circular path of radius 4.5 cm in a magnetic field with $B = 1.2$ T. Calculate (*a*) its speed, (*b*) its period of revolution, (*c*) its kinetic energy in eV, and (*d*) the potential difference through which it would have to be accelerated to achieve this energy.

14. A physicist is designing a cyclotron to accelerate protons to $0.100c$. The magnet used will produce a field of 1.40 T. Calculate (*a*) the radius of the cyclotron and (*b*) the corresponding oscillator frequency. Relativity considerations are not significant.

15. In a nuclear experiment, a proton with kinetic energy K_p moves in a uniform magnetic field in a circular path. What energy must (*a*) an alpha particle and (*b*) a deuteron have if they are to circulate in the same orbit? (For a deuteron, $q = +e$, $m = 2.0$ u; for an alpha particle, $q = +2e$, $m = 4.0$ u.)

16. A proton, a deuteron, and an alpha particle, accelerated through the same potential difference ΔV, enter a region of uniform magnetic field, moving at right angles to $\vec{\mathbf{B}}$. (*a*) Find their kinetic energies. If the radius of the proton's circular path is r_p, what are the radii of (*b*) the deuteron and (*c*) the alpha particle paths, in terms of r_p?

17. A proton, a deuteron, and an alpha particle with the same kinetic energy enter a region of uniform magnetic field, moving at right angles to $\vec{\mathbf{B}}$. The proton moves in a circle of radius r_p. In terms of r_p, what are the radii of (*a*) the deuteron path and (*b*) the alpha particle path?

18. A deuteron in a cyclotron is moving in a magnetic field with an orbit radius of 50 cm. Because of a grazing collision with a target, the deuteron breaks up, with a negligible loss of kinetic energy, into a proton and a neutron. Discuss the subsequent motions of each. Assume that the deuteron energy is shared equally by the proton and neutron at breakup.

19. (*a*) What speed would a proton need to circle the Earth at the equator, if the Earth's magnetic field is everywhere horizontal there and directed along longitudinal lines? Relativistic effects must be taken into account. Take the magnitude of the Earth's magnetic field to be 41 μT at the equator. (*b*) Draw the velocity and magnetic field vectors corresponding to this situation.

20. Compute the radius of the path of a 10.0-MeV electron moving perpendicular to a uniform 2.20-T magnetic field. Use both the (*a*) classical and (*b*) relativistic formulas. (*c*) Calculate the true period of the circular motion. Is the result independent of the speed of the electron?

21. Ionization measurements show that a particular nuclear particle carries a double charge ($=2e$) and is moving with a speed of $0.710c$. It follows a circular path of radius 4.72 m in a magnetic field of 1.33 T. Find the mass of the particle and identify it.

22. The proton synchrotron at Fermilab accelerates protons to a kinetic energy of 950 GeV. At this energy, calculate (*a*) the speed, expressed as a fraction of the speed of light; and (*b*) the magnetic field at the proton orbit that has a radius of curvature of 750 m. (The proton has a rest energy of 938 MeV; it is necessary to use relativistic formulas here.)

23. Estimate the total path length traveled by a deuteron in a cyclotron during the acceleration process. Assume an accelerating potential between the dees of 80 kV, a dee radius of 53 cm, and an oscillator frequency of 12 MHz.

24. Consider a particle of mass m and charge q moving in the xy plane under the influence of a uniform magnetic field $\vec{\mathbf{B}}$ pointing in the $+z$ direction. Write expressions for the coordinates $x(t)$ and $y(t)$ of the particle as functions of time t, assuming that the particle moves in a circle of radius R centered at the origin of coordinates.

32-4 The Hall Effect

25. A metal strip 6.5 cm long, 0.88 cm wide, and 0.76 mm thick moves with constant speed v through a magnetic field $B = 1.2$ mT perpendicular to the strip, as shown in Fig. 32-34. A potential difference of 3.9 μV is measured between points x and y across the strip. Calculate the speed v.

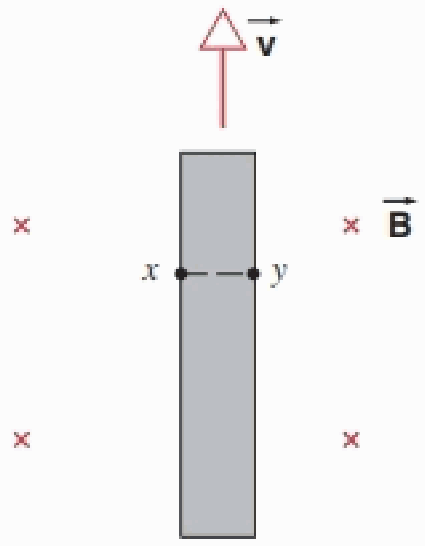


FIGURE 32-34. Exercise 25.

26. In a Hall effect experiment, a current of 3.2 A lengthwise in a conductor 1.2 cm wide, 4.0 cm long, and 9.5 μm thick produces a transverse Hall voltage (across the width) of 40 μV when a magnetic field of 1.4 T acts perpendicular to the thin conductor. From these data, find (*a*) the drift velocity of the charge carriers and (*b*) the number density of charge carriers. From Table 32-2, identify the conductor. (*c*) Show on a diagram the polarity of the Hall voltage with a given current and magnetic field direction, assuming that the charge carriers are (negative) electrons.

27. Show that, in terms of the Hall electric field E_H and the current density j, the number of charge carriers per unit volume is given by

$$n = \frac{jB}{eE_H}.$$

28. (*a*) Show that the ratio of the Hall electric field E_H to the electric field E_c responsible for the current is

$$\frac{E_H}{E_c} = \frac{B}{ne\rho},$$

where ρ is the resistivity of the material. (*b*) Compute the ratio numerically for Sample Problem 32-3. See Table 29-1.

32-5 The Magnetic Force on a Current-Carrying Wire

29. A wire of length 62.0 cm and mass 13.0 g is suspended by a pair of flexible leads in a magnetic field of 440 mT. Find the magnitude and direction of the current in the wire required to remove the tension in the supporting leads. See Fig. 32-35.

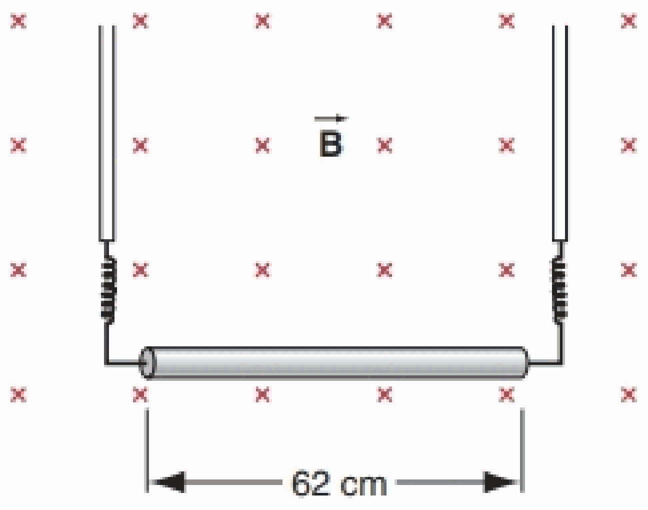


FIGURE 32-35. Exercise 29.

30. A horizontal conductor in a power line carries a current of 5.12 kA from south to north. The Earth's magnetic field in the vicinity of the line is 58.0 μT and is directed toward the north and inclined downward at 70.0° to the horizontal. Find the magnitude and direction of the magnetic force on 100 m of the conductor due to the Earth's field.

31. Consider the possibility of a new design for an electric train. The engine is driven by the force due to the vertical component of the Earth's magnetic field on a conducting axle. Current is passed down one rail, into a conducting wheel, through the axle, through another conducting wheel, and then back to the source via the other rail. (*a*) What current is needed to provide a modest 10-kN force? Take the vertical component of the Earth's field to be 10 μT and the length of the axle to be 3.0 m. (*b*) How much power would be lost for each ohm of resistance in the rails? (*c*) Is such a train totally unrealistic or just marginally unrealistic?

32. A metal wire of mass m slides without friction on two horizontal rails spaced a distance d apart, as in Fig. 32-36. The track lies in a vertical uniform magnetic field $\vec{\mathbf{B}}$. A constant current i flows from generator G along one rail, across the wire, and back down the other rail. Find the velocity (speed and direction) of the wire as a function of time, assuming it to be at rest at $t = 0$.

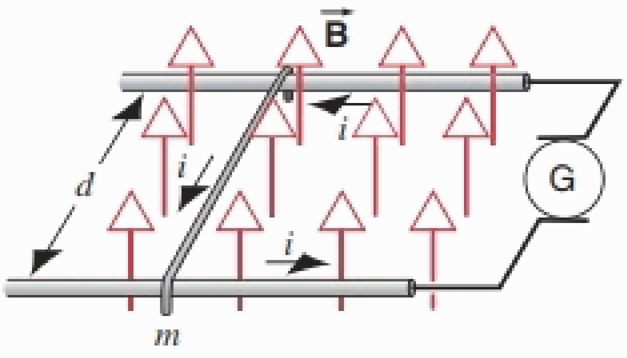


FIGURE 32-36. Exercise 32.

33. A long, rigid conductor, lying along the x axis, carries a current of 5.0 A in the $-x$ direction. A magnetic field $\vec{\mathbf{B}}$ is present, given by $\vec{\mathbf{B}} = (3 \text{ mT})\hat{\mathbf{i}} + (8 \text{ mT/m}^2)x^2\hat{\mathbf{j}}$. Calculate the force on the 2.0-m segment of the conductor that lies between $x = 1.2$ m and $x = 3.2$ m.

34. A 1.15-kg copper rod rests on two horizontal rails 95.0 cm apart and carries a current of 53.2 A from one rail to the other. The coefficient of static friction is 0.58. Find the smallest magnetic field (not necessarily vertical) that would cause the bar to slide.

32-6 The Torque on a Current Loop

35. A single-turn current loop, carrying a current of 4.00 A, is in the shape of a right triangle with sides 50 cm, 120 cm, and 130 cm. The loop is in a uniform magnetic field of magnitude 75.0 mT whose direction is parallel to the current in the 130-cm side of the loop. (*a*) Find the magnetic force on each of the three sides of the loop. (*b*) Show that the total magnetic force on the loop is zero.

36. Figure 32-37 shows a rectangular, 20-turn loop of wire that is 12 cm by 5.0 cm. It carries a current of 0.10 A and is hinged at one side. It is mounted with its plane at an angle of 33° to the direction of a uniform magnetic field of 0.50 T. Calculate the torque about the hinge line acting on the loop.

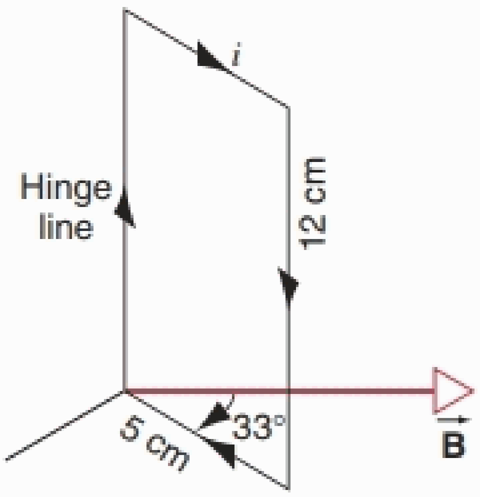


FIGURE 32-37. Exercise 36.

37. A stationary, circular wall clock has a face with a radius of 15 cm. Six turns of wire are wound around its perimeter; the wire carries a current 2.0 A in the clockwise direction. The clock is located where there is a constant, uniform external magnetic field of 70 mT (but the clock still keeps perfect time). At exactly 1:00 P.M., the hour hand of the clock points in the direction of the external magnetic field. (*a*) After how many minutes will the minute hand point in the direction of the torque on the winding due to the magnetic field? (*b*) What is the magnitude of this torque?

PROBLEMS

1. An electron in a uniform magnetic field has a velocity $\vec{\mathbf{v}} = (40\text{ km/s})\hat{\mathbf{i}} + (35\text{ km/s})\hat{\mathbf{j}}$. It experiences a force $\vec{\mathbf{F}} = (-4.2\text{ fN})\hat{\mathbf{i}} + (4.8\text{ fN})\hat{\mathbf{j}}$. If $B_x = 0$, calculate the magnetic field.

2. An electron has an initial velocity $(12.0\text{ km/s})\hat{\mathbf{j}} + (15.0\text{ km/s})\hat{\mathbf{k}}$ and a constant acceleration of $(2.00 \times 10^{12}\text{ m/s}^2)\hat{\mathbf{i}}$ in a region in which uniform electric and magnetic fields are present. If $\vec{\mathbf{B}} = (400\ \mu\text{T})\hat{\mathbf{i}}$, find the electric field $\vec{\mathbf{E}}$.

3. The electrons in the beam of a television tube have a kinetic energy of 12.0 keV. The tube is oriented so that the electrons move horizontally from magnetic south to magnetic north. The vertical component of the Earth's magnetic field points down and has a magnitude of 55.0 μT. (*a*) In what direction will the beam deflect? (*b*) What is the acceleration of a given electron due to the magnetic field? (*c*) How far will the beam deflect in moving 20.0 cm through the television tube?

4. A beam of electrons whose kinetic energy is K emerges from a thin-foil "window" at the end of an accelerator tube. There is a metal plate a distance d from this window and at right angles to the direction of the emerging beam (see Fig. 32-38). (*a*) Show that we can prevent the beam from hitting the plate if we apply a magnetic field B such that

$$B \geq \sqrt{\frac{2mK}{e^2 d^2}},$$

in which m and e are the electron mass and charge. (*b*) How should B be oriented?

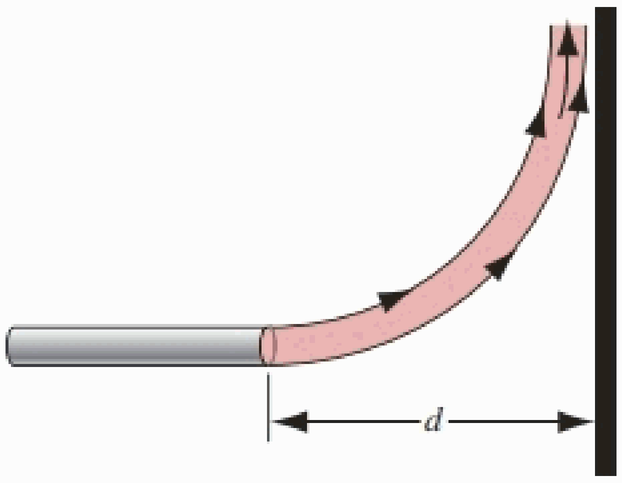


FIGURE 32-38. Problem 4.

5. Bainbridge's mass spectrometer, shown in Fig. 32-39, separates ions having the same velocity. The ions, after entering through slits S_1 and S_2, pass through a velocity selector composed of an electric field produced by the charged plates P and P′, and a magnetic field $\vec{\mathbf{B}}$ perpendicular to the electric field and the ion path. Those ions that pass undeviated through the crossed $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ fields enter into a region where a second magnetic field $\vec{\mathbf{B}}'$ exists, and are bent into circular paths. A photographic plate registers their arrival. Show that $q/m = E/rBB'$, where r is the radius of the circular orbit.

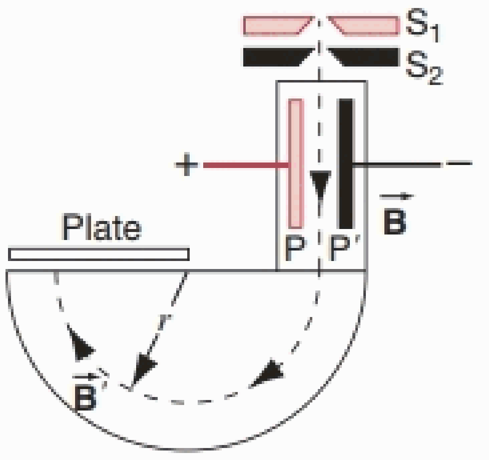


FIGURE 32-39. Problem 5.

6. Figure 32-40 shows an arrangement used to measure the masses of ions. An ion of mass m and charge $+q$ is produced essentially at rest in source S, a chamber in which a gas discharge is taking place. The ion is accelerated by potential difference ΔV and allowed to enter a magnetic field $\vec{\mathbf{B}}$. In the field it moves in a semicircle, striking a photographic plate at distance x from the entry slit. Show that the ion mass m is given by

$$m = \frac{B^2 q}{8\Delta V} x^2.$$

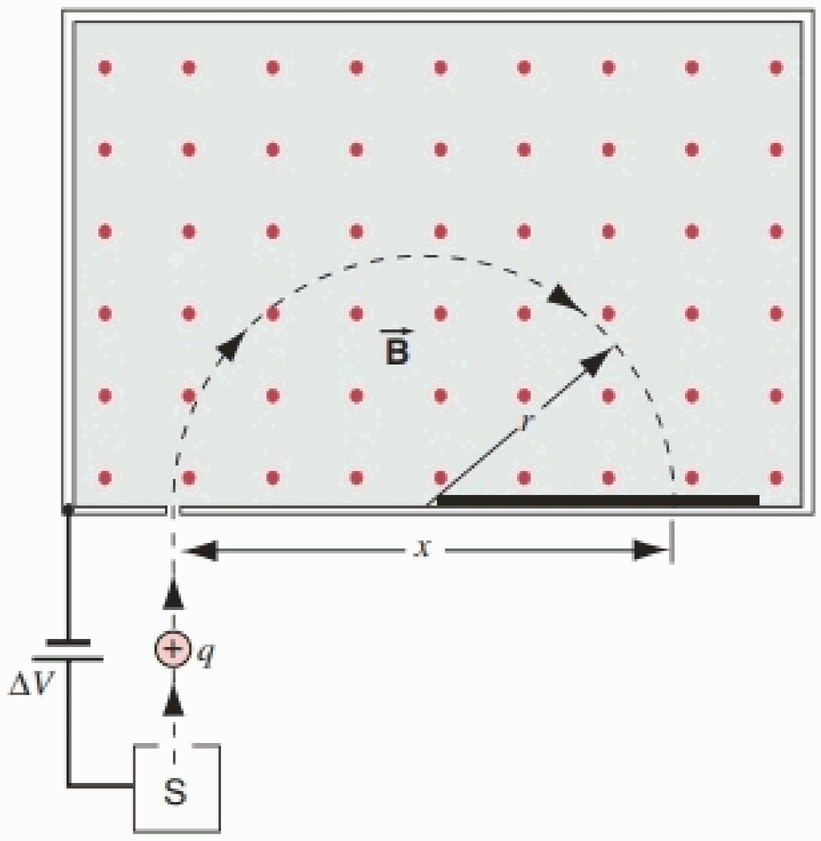


FIGURE 32-40. Problem 6.

7. Two types of singly ionized atoms having the same charge q and mass differing by a small amount Δm are introduced into the mass spectrometer described in Problem 6. (*a*) Calculate the difference in mass in terms of ΔV, q, m (of either), B, and the distance Δx between the spots on the photographic plate. (*b*) Calculate Δx for a beam of singly ionized chlorine atoms of masses 35.0 u and 37.0 u if $\Delta V = 7.33$ kV and $B =$ 520 mT.

8. In a mass spectrometer (see Problem 6) used for commercial purposes, uranium ions of mass 238 u and charge $+2e$ are separated from related species. The ions are first accelerated through a potential difference of 105 kV and then pass into a magnetic field, where they travel a 180° arc of radius 97.3 cm. They are then collected in a cup after passing through a slit of width 1.20 mm and a height of 1.14 cm. (*a*) What is the magnitude of the (perpendicular) magnetic field in the separator? If the machine is designed to separate out 90.0 mg of material per hour, calculate (*b*) the current of the desired ions in the machine and (*c*) the internal energy dissipated in the cup in 1.00 h.

9. A neutral particle is at rest in a uniform magnetic field of magnitude B. At time $t = 0$ it decays into two charged particles each of mass m. (a) If the charge of one of the particles is $+q$, what is the charge of the other? (b) The two particles move off in separate paths, both of which lie in the plane perpendicular to $\vec{\mathbf{B}}$. At a later time the particles collide. Express the time from decay until collision in terms of m, B, and q.

10. In Bohr's theory of the hydrogen atom, the electron can be thought of as moving in a circular orbit of radius r about the proton. Suppose that such an atom is placed in a magnetic field, with the plane of the orbit at right angles to $\vec{\mathbf{B}}$. (a) If the electron is circulating clockwise, as viewed by an observer sighting along $\vec{\mathbf{B}}$, will the angular frequency increase or decrease? (b) What if the electron is circulating counterclockwise? Assume that the orbit radius does not change. [Hint: The centripetal force is now partially electric ($\vec{\mathbf{F}}_E$) and partially magnetic ($\vec{\mathbf{F}}_B$) in origin.] (c) Show that the change in frequency of revolution caused by the magnetic field is given approximately by

$$\Delta f = \pm \frac{Be}{4\pi m}.$$

Such frequency shifts were observed by Zeeman in 1896. (Hint: Calculate the frequency of revolution without the magnetic field and also with it. Subtract, bearing in mind that because the effect of the magnetic field is very small, some—but not all—terms containing B can be set equal to zero with little error.)

11. A 22.5-eV positron (positively charged electron) is projected into a uniform magnetic field $B = 455\ \mu\text{T}$ with its velocity vector making an angle of 65.5° with $\vec{\mathbf{B}}$. Find (a) the period, (b) the pitch p, and (c) the radius r of the helical path. See Fig. 32-41.

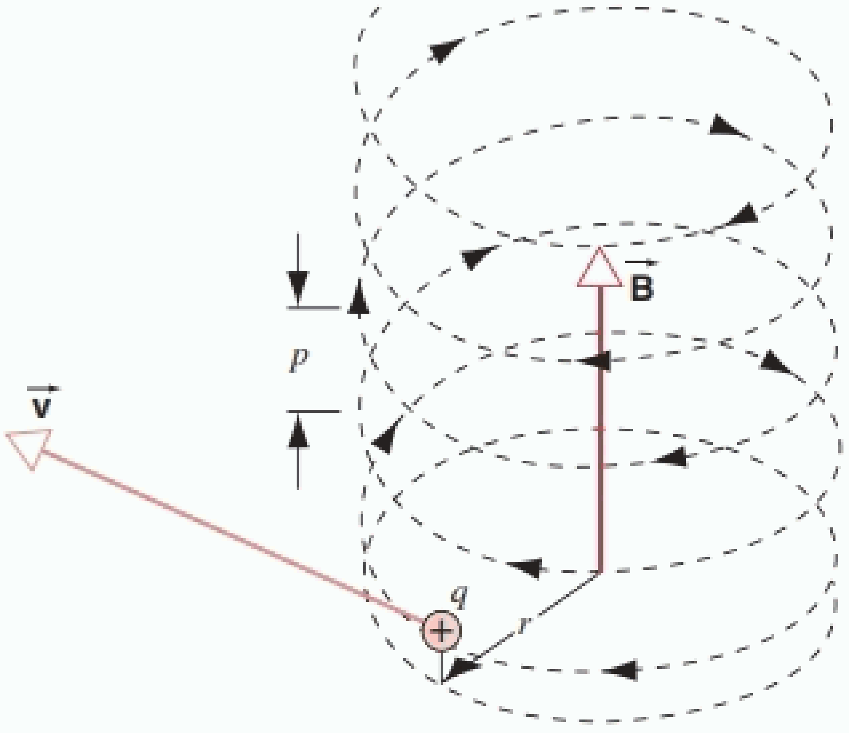


FIGURE 32-41. Problem 11.

12. Figure 32-42 shows a wire of arbitrary shape carrying a current i between points a and b. The wire lies in a plane at right angles to a uniform magnetic field $\vec{\mathbf{B}}$. Prove that the force on the wire is the same as that on a straight wire carrying a current i directly from a to b. (Hint: Replace the wire by a series of "steps" that are parallel and perpendicular to the straight line joining a and b.)

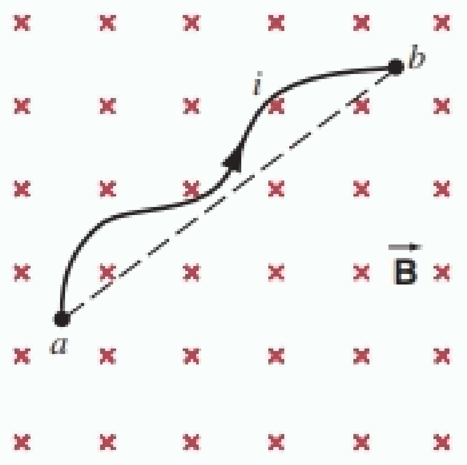


FIGURE 32-42. Problem 12.

13. Consider the particle of Exercise 24, but this time prove (rather than assume) that the particle moves in a circular path by solving Newton's second law analytically. [Hint: Solve the expression for F_y to find v_x and substitute into the expression for F_x to obtain an equation that can be solved for v_y. Do the same for v_x by substituting into the F_y equation. Finally, obtain $x(t)$ and $y(t)$ from v_x and v_y.]

14. By direct integration of

$$\vec{\mathbf{F}}_B = \oint i\, d\vec{\mathbf{L}} \times \vec{\mathbf{B}},$$

show that the net force on an *arbitrary* current loop is zero in a uniform magnetic field. (*Note: an arbitrary current loop does not need to lie in a plane!*)

15. A U-shaped wire of mass m and length L is immersed with its two ends in mercury (Fig. 32-43). The wire is in a homogeneous magnetic field $\vec{\mathbf{B}}$. If a charge—that is, a current pulse $q = \int i dt$—is sent through the wire, the wire will jump up. Calculate, from the height h that the wire reaches, the size of the charge or current pulse, assuming that the time of the current pulse is very small in comparison with the time of flight. Make use of the fact that impulse of force equals $\int F dt$, which equals mv. (Hint: Relate $\int i dt$ to $\int F dt$.) Evaluate q for $B = 0.12$ T, $m = 13$ g, $L = 20$ cm, and $h = 3.1$ m.

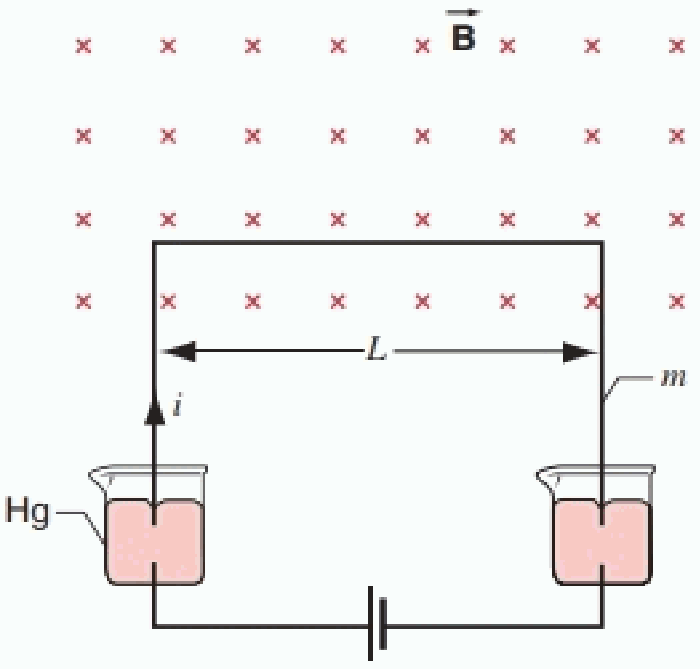


FIGURE 32-43. Problem 15.

16. Prove that Eq. 32-34 holds for closed loops of arbitrary shape and not only for rectangular loops as in Fig. 32-26. (Hint: Replace the loop of arbitrary shape by an assembly of adjacent, long, thin, approximately rectangular loops that are nearly equivalent to it as far as the distribution of current is concerned.)

17. A length L of wire carries a current i. Show that if the wire is formed into a circular coil, the maximum torque in a given magnetic field is developed when the coil has one turn only and the maximum torque has the magnitude

$$\tau = \frac{1}{4\pi} L^2 iB.$$

18. Figure 32-44 shows a wire ring of radius a at right angles to the general direction of a radially symmetric diverging magnetic field. The magnetic field at the ring is everywhere of the same magnitude B, and its direction at the ring is everywhere at an angle θ with a normal to the plane of the ring. The twisted lead wires have no effect on the problem. Find the magnitude and direction of the force the field exerts on the ring if the ring carries a current i as shown in the figure.

19. Figure 32-45 shows a wooden cylinder with a mass m = 262 g and a length L = 12.7 cm, with N = 13 turns of wire wrapped around it longitudinally, so that the plane of the wire loop contains the axis of the cylinder. What is the least current through the loop that will prevent the cylinder from rolling down a plane inclined at an angle θ to the horizontal, in the presence of a vertical, uniform magnetic field of 477 mT, if the plane of the windings is parallel to the inclined plane?

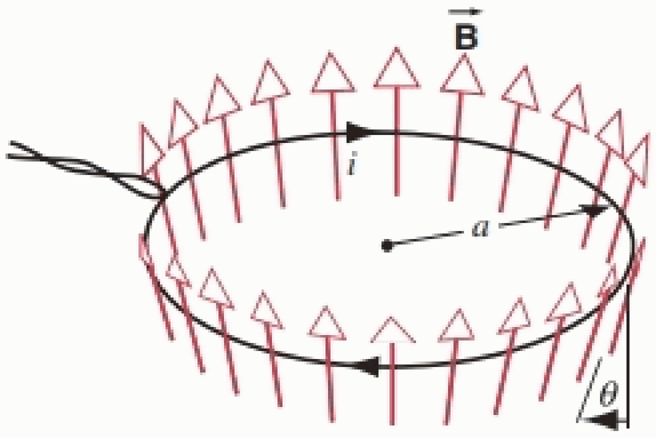


FIGURE 32-44. Problem 18.

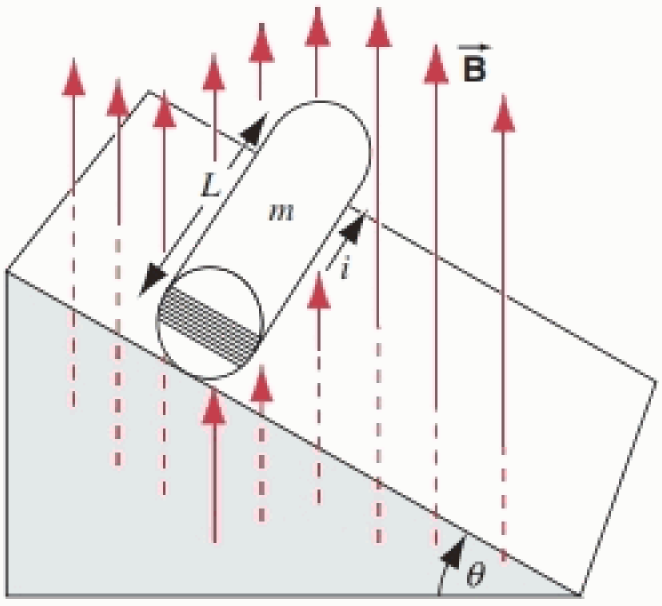


FIGURE 32-45. Problem 19.

COMPUTER PROBLEMS

1. Using the magnetic field given in Eq. 32-17 with B_0 = 0.15 T, obtain the trajectory of an alpha particle that is initially moving through the origin in the x direction with speed v_0 = 3.0×10^6 m/s (Fig. 32-17). Find the time that it takes the particle to return to its starting location and its maximum distance from the origin. Compare these values with the corresponding ones in the case of a uniform field B_0.

2. A cylindrically symmetric magnetic field in a certain region of space is given by $\vec{\mathbf{B}} = (B_0 r/a)\hat{\mathbf{k}}$, where r is the perpendicular distance from the z axis. Find the trajectory of an electron emitted from a point on the z axis with an initial velocity of $0.050c$ perpendicular to the axis. What is the maximum distance the electron will travel from the axis before returning?

CHAPTER 33

THE MAGNETIC FIELD OF A CURRENT

In the previous chapter we studied the effect of a magnetic field on a moving charge. We now turn to the source of the field itself, and in this chapter we study the magnetic field produced by moving charges, particularly currents in wires.

In analogy with our previous study of the electric fields of some simple charge distributions, we investigate in this chapter the magnetic fields produced by some simple current distributions: straight wires and circular loops. Finally, we show that the relationship between electric and magnetic fields is deeper than merely the similarity of equations; the relationship extends to the transformation of the fields into one another when charge or current distributions are viewed from different inertial frames.

33-1 THE MAGNETIC FIELD DUE TO A MOVING CHARGE

In the previous chapter we discussed the force experienced by a charged particle moving in a magnetic field. By analogy with the electric field, in which charged particles at rest are both the sources of the field and its probe, we might expect that moving electric charges, which we have already shown to be probes of the magnetic field, could also serve as sources of the field.

This expectation was first demonstrated in 1820 by Hans Christian Oersted*, who observed that, as illustrated in Fig. 33-1, when a compass is placed near a straight wire carrying a current, the compass needle aligns so that it is tangent to a circle drawn around the wire (neglecting the influence of the Earth's magnetic field on the compass). Oersted's discovery provided the first link between electricity and magnetism.

Direct experimental evidence for the magnetic field of a moving charge did not come until 1876 in an experiment by Henry Rowland.** Rowland's experiment is shown schematically in Fig. 33-2. He prepared a disk of charge (by connecting a battery to a layer of gold deposited on the surface of a disk of insulating material). By rotating the disk about its axis, he was able to produce moving charges, and he showed their magnetic effect by suspending a magnetized needle near the disk.

Our goal in this chapter is to study the magnetic interaction between two moving charges, just as Coulomb studied the electric interaction between charges at rest. Coulomb was able to measure the electrostatic force directly, and in principle we should be able to do the same—that is, to measure the magnetic force between two charges in motion. Unfortunately, the force is extremely small and very difficult to measure; in Rowland's experiment, for example, the magnetic field produced by his rotating charged disk was only 0.00001 of the Earth's field!

Despite the extremely small size of the magnetic field of a single moving charge, it is conceptually easier to begin our study of the sources of the magnetic field with a discussion of how a single moving charge produces a magnetic field. Later we will see why this approach is not practical and why it is easier for us to produce magnetic fields in the laboratory by using moving charges in the form of currents in wires.

---

*Hans Christian Oersted (1777–1851) was a Danish physicist and chemist. His discovery that a current-carrying wire can deflect a compass needle was made unexpectedly during a lecture at the University of Copenhagen. The unit for magnetic field intensity (H), the oersted, is named in his honor.

**Henry Rowland (1848–1901) was a U.S. physicist who is today most remembered for his pioneering work in developing diffraction gratings, which he used for precise optical spectroscopy including measurements of the wavelengths of the solar spectrum. Rowland served as the first president of the American Physical Society.

Let us therefore conduct a "thought experiment" in which we project a single charge q with velocity $\vec{\mathbf{v}}$ and detect the field with a suspended magnetic needle that is free to align in any direction. To avoid problems with relativity, we will need to keep the speed of the particle small (compared with the speed of light) in our reference frame. We set up the experiment in a region in which the Earth's magnetic field is negligible. (It is not necessary to journey to outer space to find this region; we can use current-carrying coils in our laboratories to create fields that cancel the Earth's field.) Figure 33-3*a* shows the outcome of some measurements of the magnetic field at different locations. The moving charge sets up a magnetic field $\vec{\mathbf{B}}$, and the needle indicates the direction of the field at any location. In principle we could also determine the magnitude of the field, such as by measuring the force on a second moving charged particle, as we described in Section 32-2.

If we could actually perform these experiments, we would discover some properties of the magnetic field due to a moving charge:

1. The field strength is directly proportional to the magnitude of the velocity v and also to the charge q.

2. If $\vec{\mathbf{v}}$ reverses direction or q changes sign, the direction of $\vec{\mathbf{B}}$ is reversed.

3. The field is zero at points along the direction of $\vec{\mathbf{v}}$ (forward as well as backward). In other directions relative to $\vec{\mathbf{v}}$, as shown in Fig. 33-3*b*, the field varies as $\sin \boldsymbol{\phi}$.

4. $\vec{\mathbf{B}}$ is tangent to circles drawn about $\vec{\mathbf{v}}$ in planes perpendicular to $\vec{\mathbf{v}}$, with the direction of $\vec{\mathbf{B}}$ determined by a right-hand rule (point your thumb in the direction of $\vec{\mathbf{v}}$, and your fingers will curl in the direction of $\vec{\mathbf{B}}$). On any given circle, the magnitude of $\vec{\mathbf{B}}$ is the same at all points.

5. At points on a line perpendicular to the direction of motion of q (as in Fig. 33-3*b*) or equivalently on circles of

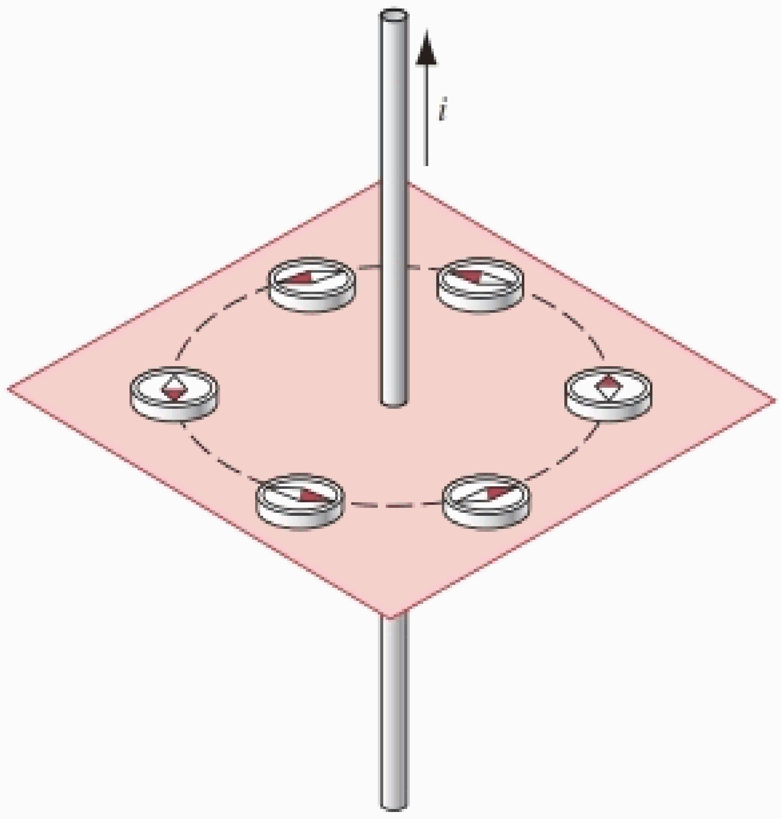


FIGURE 33-1. Oersted's experiment. The direction of the compass needle is always perpendicular to the direction of the current in the wire.

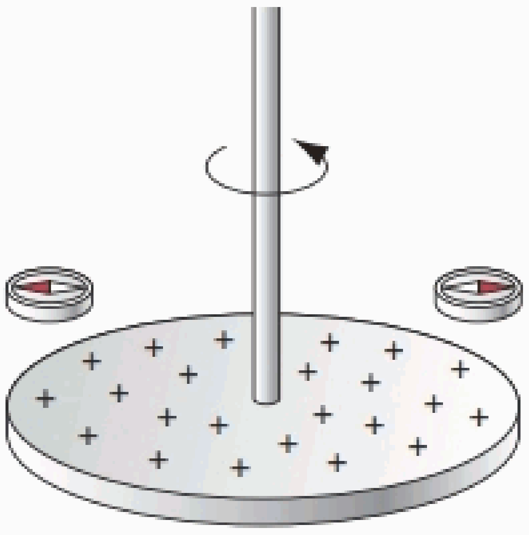


FIGURE 33-2. Schematic diagram of Rowland's experiment. The moving charges on the gold surface of the rotating disk produce a magnetic field that deflects the compass needle. In practice, the deflection is very small and requires apparatus far more sensitive than a compass to detect.

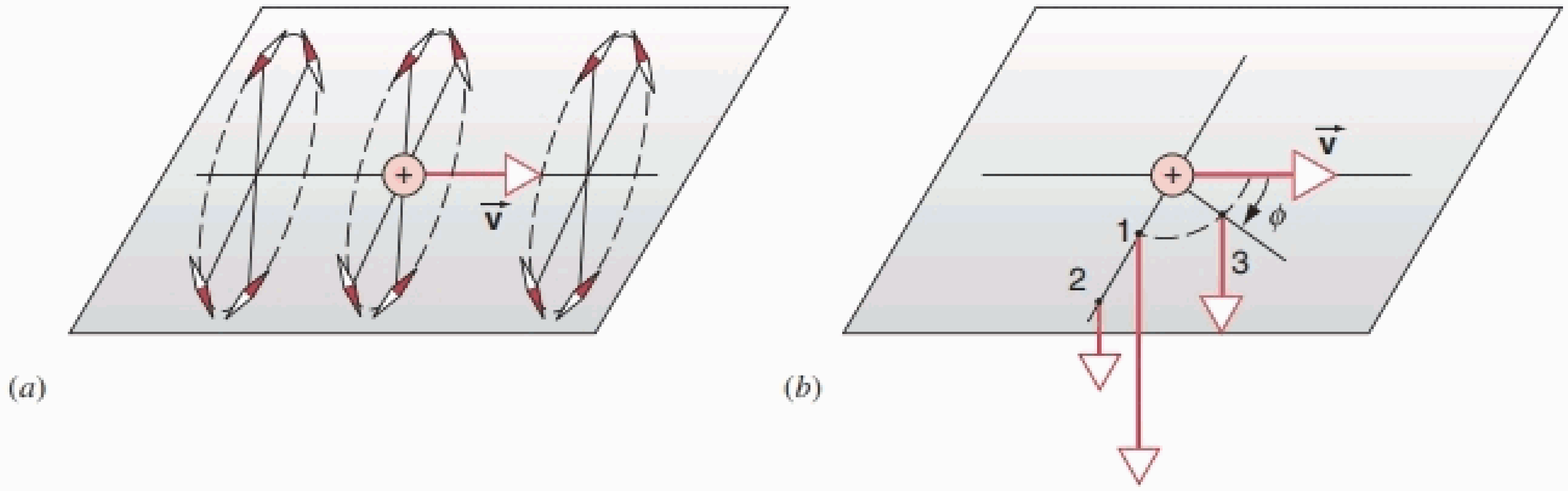


FIGURE 33-3. (*a*) A freely suspended compass needle indicates the direction of the magnetic field at various locations due to a moving electric charge. (*b*) The field at point 2 is 1/4 of the field at point 1, because point 2 is twice as far from the charge. Point 3 is the same distance from the charge as point 1, but the field at point 3 is smaller than the field at point 1 by the factor $\sin\phi$.

increasing radius drawn around the line of motion, we find that the magnitude of $\vec{\mathbf{B}}$ decreases like $1/r^2$, where r is the distance from q to the observation point.

The simplest way of defining $\vec{\mathbf{B}}$ that is consistent with these observations is illustrated in Fig. 33-4. At an arbitrary point P (the point at which we wish to find the magnetic field), $\vec{\mathbf{B}}$ is perpendicular to the plane determined by $\vec{\mathbf{v}}$ and $\vec{\mathbf{r}}$ (the vector that locates P relative to q). We know from our observations that the magnitude of $\vec{\mathbf{B}}$ is directly proportional to q, v, and $\sin\phi$, and inversely proportional to r^2:

$$B \propto \frac{qv\sin\phi}{r^2}. \tag{33-1}$$

The direction of $\vec{\mathbf{B}}$ relative to $\vec{\mathbf{v}}$ and $\vec{\mathbf{r}}$ reminds us of the rule for finding a vector (cross) product. In the expression $\vec{\mathbf{c}} = \vec{\mathbf{a}} \times \vec{\mathbf{b}}$, the vector $\vec{\mathbf{c}}$ is perpendicular to the plane containing $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$. We can therefore write Eq. 33-1 in vector form as

$$\vec{\mathbf{B}} = K\frac{q\vec{\mathbf{v}} \times \hat{\mathbf{r}}}{r^2}, \tag{33-2}$$

where K is a constant of proportionality to be determined. Here $\hat{\mathbf{r}}$ is the unit vector in the direction of $\vec{\mathbf{r}}$. (You may wish to review Section 25-4, where we used a similar unit vector notation to express Coulomb's law.) Because $\hat{\mathbf{r}} = \vec{\mathbf{r}}/r$, we can write Eq. 33-2 as

$$\vec{\mathbf{B}} = K\frac{q\vec{\mathbf{v}} \times \vec{\mathbf{r}}}{r^3}. \tag{33-3}$$

Even though there is a factor of r^3 in the denominator, the field varies as $1/r^2$, because there is also a factor of r in the numerator.

All that remains to obtain the complete expression for the magnetic field of a moving charge is to determine the constant of proportionality in Eq. 33-3, just as we inserted the constant $1/4\pi\epsilon_0$ into Coulomb's law. The constants in the magnetic and electric field equations are not independent quantities; they are related by the speed of light, as we discuss in Chapter 38. Since the speed of light is a defined quantity, our choices are either (1) to use the electric force law (Coulomb's law) to define the electric constant and then use the value of the speed of light to obtain the magnetic constant, or (2) to use a magnetic force law (the force between current-carrying wires, which we discuss later in this chapter) to define the magnetic constant and then use the value of the speed of light to obtain the electric constant. Because the magnetic force between current-carrying wires can be measured more precisely than the electric force between charges, we choose the second method.

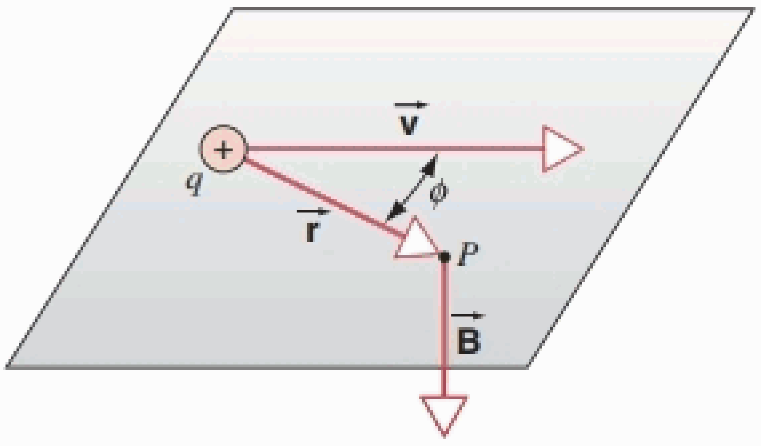


FIGURE 33-4. The magnetic field at point P due to a moving charge is perpendicular to the plane containing $\vec{\mathbf{v}}$ and $\vec{\mathbf{r}}$.

The proportionality constant K in Eq. 33-3 is defined in SI units to have the exact value 10^{-7} tesla·meter/ampere (T·m/A). However, as was the case in electrostatics, we find it convenient to write the constant in a different form:

$$K = \frac{\mu_0}{4\pi} = 10^{-7}\ \mathrm{T\cdot m/A}.$$

The constant μ_0 has historically been known as the *permeability constant,* but we will refer to it simply as the *magnetic constant.* It has the exact value

$$\mu_0 = 4\pi \times 10^{-7}\ \mathrm{T\cdot m/A}.$$

The magnetic constant μ_0 plays a role in calculating magnetic fields similar to that of the electric constant ϵ_0 in calculating electric fields. The two constants are related through the speed of light: $c = (\epsilon_0\mu_0)^{-1/2}$. Defining μ_0 and c then determines ϵ_0 exactly.

We are now able to write the complete expression for the magnetic field due to a moving charge:

$$\vec{\mathbf{B}} = \frac{\mu_0}{4\pi}\frac{q\vec{\mathbf{v}} \times \hat{\mathbf{r}}}{r^2} = \frac{\mu_0}{4\pi}\frac{q\vec{\mathbf{v}} \times \vec{\mathbf{r}}}{r^3}. \tag{33-4}$$

We can write the magnitude of $\vec{\mathbf{B}}$ as

$$B = \frac{\mu_0}{4\pi}\frac{|q|v\sin\phi}{r^2}, \tag{33-5}$$

where ϕ is the angle between $\vec{\mathbf{v}}$ and $\vec{\mathbf{r}}$.

Based on these thought experiments, we have been able to learn several properties of the magnetic field produced by a moving charge, including the important geometrical relationship between the direction of the velocity and the direction of the field. It will now be trivial to transfer this relationship to the more useful case of the magnetic field produced by currents in wires.

Why is the magnetic field specified by Eqs. 33-4 or 33-5 not especially useful? When we studied electric fields, our interest was in the steady electric field produced by charges whose locations did not change. We called this subject *electrostatics.* Now we are interested in *magnetostatics*—the production of steady magnetic fields by charges whose motion does not change. The single moving charge of Fig. 33-4 does not satisfy this criterion; an instant after the situation illustrated in the figure, the charge is at a different location with respect to point P, and there is no charge at all at its previous location. An instant later there will be a different magnetic field at P. To maintain a steady magnetic field at P due to a moving charge at the exact location shown, it would be necessary to arrange to destroy the charge as soon as it left that location and to inject a new charge into that location with the same velocity, a highly improbable scenario. On the other hand, a steady current accomplishes exactly what we want: an unchanging motion of charges that produces a steady magnetic field. In the next section we adapt Eqs. 33-4 and 33-5 to the case of steady currents.

SAMPLE PROBLEM 33-1. An alpha particle ($q = +2e$) is moving in the positive x direction with a speed of $0.0050c = 1.50 \times 10^6$ m/s. When the particle is at the origin, find the magnetic field at (*a*) P_1: $x = 0$, $y = 0$, $z = +2.0$ cm; (*b*) P_2: $x = 0$, $y = +2.0$ cm, $z = 0$; (*c*) P_3: $x = +1.0$ cm, $y = +1.0$ cm, $z = +1.0$ cm.

Solution (*a*) In Fig. 33-5, $\vec{\mathbf{r}}_1$ is in the $+z$ direction (pointing from q to P_1). The length of $\vec{\mathbf{r}}_1$ is the distance from the origin to P_1, or 2.0 cm. The vectors $\vec{\mathbf{v}}$ and $\vec{\mathbf{r}}_1$ are in the xz plane; $\vec{\mathbf{B}}$ must be perpendicular to that plane, or in the positive or negative y direction. The direction of $\vec{\mathbf{v}} \times \vec{\mathbf{r}}_1$ determines that $\vec{\mathbf{B}}$ is in the $-y$ direction. The magnitude of $\vec{\mathbf{B}}$ is given by Eq. 33-5:

$$B = \frac{\mu_0}{4\pi}\frac{qv\sin\phi}{r_1^2}$$

$$= (10^{-7}\,\mathrm{T\cdot m/A})\,\frac{(2)(1.60 \times 10^{-19}\,\mathrm{C})(1.50 \times 10^6\,\mathrm{m/s})(\sin 90^\circ)}{(0.020\,\mathrm{m})^2}$$

$$= 1.2 \times 10^{-16}\,\mathrm{T}.$$

(*b*) At P_2, $\vec{\mathbf{r}}_2$ is in the $+y$ direction, and so $\vec{\mathbf{v}} \times \vec{\mathbf{r}}_2$ determines that $\vec{\mathbf{B}}$ is in the $+z$ direction. Because the distance from q to P_2 is the same as the distance from q to P_1, the magnitude of $\vec{\mathbf{B}}$ at P_2 is the same as we found in part (*a*) for P_1. In fact, $\vec{\mathbf{B}}$ has this same magnitude at all points on the circle of radius $r = 2.0$ cm drawn about q in the yz plane.

(*c*) At P_3, $r_3 = \sqrt{x^2 + y^2 + z^2} = 1.73$ cm. The vector $\vec{\mathbf{r}}_3$ from q to P_3 and the vector $\vec{\mathbf{v}}$ form a plane that makes an angle of 45° with the y and z axes (Fig. 33-5). The angle ϕ between $\vec{\mathbf{v}}$ and $\vec{\mathbf{r}}_3$ is 54.7°, as you should show. The magnitude of $\vec{\mathbf{B}}$ at P_3 is

$$B = \frac{\mu_0}{4\pi}\frac{qv\sin\phi}{r_3^2}$$

$$= (10^{-7}\,\mathrm{T\cdot m/A})$$

$$\times \frac{(2)(1.60 \times 10^{-19}\,\mathrm{C})(1.50 \times 10^6\,\mathrm{m/s})(\sin 54.7^\circ)}{(0.0173\,\mathrm{m})^2}$$

$$= 1.3 \times 10^{-16}\,\mathrm{T}.$$

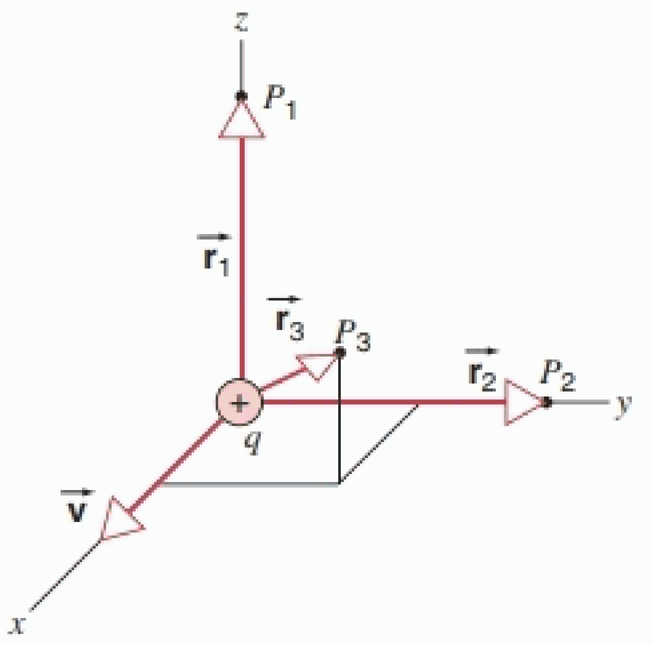


FIGURE 33-5. Sample Problem 33-1.

$\vec{\mathbf{B}}$ is perpendicular to the plane of $\vec{\mathbf{r}}_3$ and $\vec{\mathbf{v}}$. Its Cartesian components B_y and B_z are equal in magnitude, but B_y is negative and B_z is positive. Equivalently, $\vec{\mathbf{B}}$ is tangent to the circle drawn about the x axis but centered at $x = 1.0$ cm, not $x = 0$. This circle has radius $\sqrt{y^2 + z^2} = 1.41$ cm. Note in this case that, in contrast to parts (*a*) and (*b*), r_3 is not the radius of the circle to which $\vec{\mathbf{B}}$ is tangent.

The magnetic fields we found in this problem are very small, about 12 orders of magnitude smaller than the Earth's field. You can see from this calculation why it is hopeless to try to measure the force between individual charged particles in motion. However, the fields due to individual particles at distances r corresponding to atomic dimensions (10^{-10} m) are of order 1 T, which can certainly produce measurable effects. On the scale of atoms, the magnetic force between moving charged particles often has observable consequences, as we discuss in Chapter 35.

33-2 THE MAGNETIC FIELD OF A CURRENT

In the laboratory we produce magnetic fields using current-carrying wires rather than the motion of individual charges. In this section we extend the results of the previous section to permit us to find the magnetic field due to a current. Our strategy is first to find the field due to the current in a short element of the wire, and then to use integration methods to find the field due to the current in the entire wire.

This method is similar to the one we used in Section 26-4 to find the electric field due to a continuous charge distribution. In Chapter 26, we began with the electric field due to a point charge, which we can write as $\vec{\mathbf{E}} = (q/4\pi\epsilon_0 r^2)\hat{\mathbf{r}}$. To find the electric field due to a continuous charge distribution, we imagined the object to be composed of infinitesimal elements of charge dq, each of which could be treated as a point charge in calculating its contribution $d\vec{\mathbf{E}}$ to the electric field: $d\vec{\mathbf{E}} = (dq/4\pi\epsilon_0 r^2)\hat{\mathbf{r}}$. The total electric field is found by adding the contributions of all the charge elements: $\vec{\mathbf{E}} = \int d\vec{\mathbf{E}}$, which is a shorthand way of representing the total field of each of the vector components: $E_x = \int dE_x$ and so forth.

How do we represent an increment of current in the analogous calculation of the magnetic field? For a clue about how to proceed, we can review the relationship from the previous chapter between the magnetic force exerted on an individual moving charge, $\vec{\mathbf{F}}_B = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}$ (Eq. 32-3), and the magnetic force exerted on a current element, $d\vec{\mathbf{F}}_B = i\,d\vec{\mathbf{L}} \times \vec{\mathbf{B}}$ (Eq. 32-29). In this equation, $d\vec{\mathbf{L}}$ is a vector whose length is equal to the length of the element of the wire and whose direction is the direction of the current in that element. That is, we can go from an expression describing the force on an individual moving charge to one describing the force on a current element by replacing $q\vec{\mathbf{v}}$ by $i\,d\vec{\mathbf{L}}$.

We can modify Eq. 33-4 in exactly the same way. We seek the contribution $d\vec{\mathbf{B}}$ to the magnetic field due to a current element, which we can consider to be represented by a charge element dq moving with velocity $\vec{\mathbf{v}}$:

$$d\vec{\mathbf{B}} = \frac{\mu_0}{4\pi}\frac{dq\vec{\mathbf{v}} \times \hat{\mathbf{r}}}{r^2}. \qquad (33\text{-}6)$$

We can write the velocity as $\vec{\mathbf{v}} = d\vec{\mathbf{s}}/dt$, so that the charge dq moves through the displacement $d\vec{\mathbf{s}}$ in the time interval dt. We now have

$$dq\,\vec{\mathbf{v}} = dq\,\frac{d\vec{\mathbf{s}}}{dt} = \frac{dq}{dt}\,d\vec{\mathbf{s}} = i\,d\vec{\mathbf{s}}. \qquad (33\text{-}7)$$

Substituting Eq. 33-7 into Eq. 33-6, we obtain

$$d\vec{\mathbf{B}} = \frac{\mu_0}{4\pi}\frac{i\,d\vec{\mathbf{s}} \times \hat{\mathbf{r}}}{r^2} = \frac{\mu_0}{4\pi}\frac{i\,d\vec{\mathbf{s}} \times \vec{\mathbf{r}}}{r^3}. \qquad (33\text{-}8)$$

This expression is known as the *Biot–Savart law.* The direction of $d\vec{\mathbf{B}}$ is the same as the direction of $d\vec{\mathbf{s}} \times \vec{\mathbf{r}}$. The magnitude of the field element $d\vec{\mathbf{B}}$ is

$$dB = \frac{\mu_0}{4\pi}\frac{i\,ds\sin\phi}{r^2}, \qquad (33\text{-}9)$$

where ϕ is the angle between the vector $d\vec{\mathbf{s}}$, which indicates the direction of the current, and the vector $\vec{\mathbf{r}}$ from the current element to the observation point P. Figure 33-6 shows the vector relationships; compare with Fig. 33-4 and note how similar the two figures are.

To find the total field $\vec{\mathbf{B}}$ due to the entire current distribution, we must integrate over all current elements $i\,d\vec{\mathbf{s}}$:

$$\vec{\mathbf{B}} = \int d\vec{\mathbf{B}} = \frac{\mu_0}{4\pi}\int\frac{i\,d\vec{\mathbf{s}} \times \hat{\mathbf{r}}}{r^2} = \frac{\mu_0}{4\pi}\int\frac{i\,d\vec{\mathbf{s}} \times \vec{\mathbf{r}}}{r^3}. \qquad (33\text{-}10)$$

Just as we did in Chapter 26 for electric fields, we will in general have to take into account in computing this integral that not all of the elements of $d\vec{\mathbf{B}}$ are in the same direction; see Section 26-4.

We now consider how to apply the Biot–Savart law to calculate the magnetic fields of some current-carrying wires of different shapes.

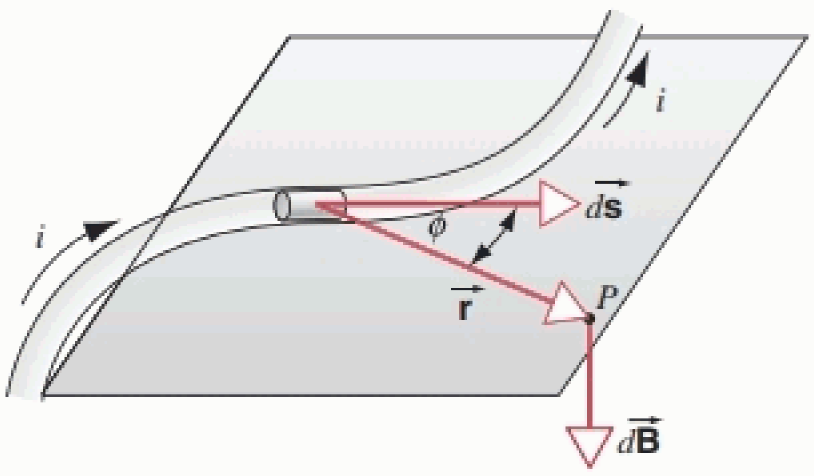


FIGURE 33-6. The magnetic field $d\vec{\mathbf{B}}$ produced by an element of a current-carrying wire. In analogy with Fig. 33-4, the field is perpendicular to the plane containing $d\vec{\mathbf{s}}$ and $\vec{\mathbf{r}}$.

A Straight Wire Segment

We illustrate the use of the Biot–Savart law by applying it to the calculation of the magnetic field due to a current i in a straight wire segment of length L. Figure 33-7 shows the geometry. The wire lies along the z axis, and we want to find $\vec{\mathbf{B}}$ at point P on the y axis, a distance d from the wire. The center of the wire is at the origin, so P is on the perpendicular bisector of the wire. The first step in the calculation is to choose an arbitrary element of the wire $i\,d\vec{\mathbf{s}}$, which is located at coordinate z relative to the origin. The contribution $d\vec{\mathbf{B}}$ of this element to the field at P is given by Eq. 33-8 and involves the vector cross product $d\vec{\mathbf{s}} \times \vec{\mathbf{r}}$. Using the right-hand rule we can show that, in the geometry of Fig. 33-7, $d\vec{\mathbf{s}} \times \vec{\mathbf{r}}$ is a vector that points in the negative x direction, and we see that this is true no matter where on the wire we choose the current element. Every element $i\,d\vec{\mathbf{s}}$ of the wire gives a $d\vec{\mathbf{B}}$ in the negative x direction, and therefore when we add up all the elements $d\vec{\mathbf{B}}$ we will find that the total field is in the negative x direction. Since we have now obtained the direction of $\vec{\mathbf{B}}$, we can turn to finding its magnitude using Eq. 33-9.

With $d\vec{\mathbf{s}}$ in the z direction we have $ds = dz$, and z will be our variable of integration, ranging from $-L/2$ to $+L/2$. To integrate Eq. 33-9, we must first express ϕ and r in terms of the integration variable z:

$$r = \sqrt{z^2 + d^2}$$

and

$$\sin\phi = \sin(\pi - \phi) = \frac{d}{\sqrt{z^2 + d^2}}.$$

Making these substitutions into Eq. 33-9, we obtain

$$dB = \frac{\mu_0 i}{4\pi}\frac{dz\sin\phi}{r^2} = \frac{\mu_0 i}{4\pi}\frac{d}{(z^2 + d^2)^{3/2}}\,dz. \qquad (33\text{-}11)$$

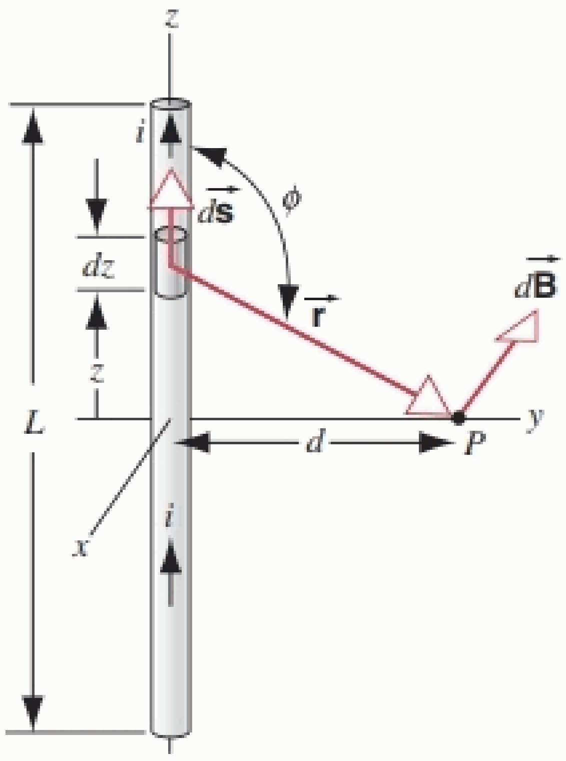


FIGURE 33-7. An element $i\,d\vec{\mathbf{s}}$ in a straight wire segment establishes at point P a field $d\vec{\mathbf{B}}$ in the negative x direction.

Carrying out the integral, we find the total field:

$$B = \frac{\mu_0 i d}{4\pi}\int_{-L/2}^{+L/2}\frac{dz}{(z^2+d^2)^{3/2}} = \frac{\mu_0 i}{4\pi d}\,\frac{z}{(z^2+d^2)^{1/2}}\Bigg|_{z=-L/2}^{z=+L/2}$$

or

$$B = \frac{\mu_0 i}{4\pi d}\,\frac{L}{(L^2/4+d^2)^{1/2}}\,. \tag{33-12}$$

In the limit of a very long wire (that is, $L \gg d$), Eq. 33-12 becomes

$$B = \frac{\mu_0 i}{2\pi d}\,. \tag{33-13}$$

This problem reminds us of its electrostatic equivalent. We derived an expression for $\vec{\mathbf{E}}$ due to a long charged rod by integration methods, using Coulomb's law (Section 26-4). We also solved the same problem using Gauss' law (Section 27-5). Later we consider a law of magnetic fields, Ampère's law, which is similar to Gauss' law in that it simplifies magnetic field calculations in cases (such as this one) that have a high degree of symmetry.

Just as we did for electric fields, we can represent the magnetic field of a current-carrying wire by magnetic field lines. Figure 33-8 shows a set of magnetic field lines for a long, straight wire. The field lines form concentric circles around the wire, as suggested by Oersted's experiment (Fig. 33-1) and as also indicated by the pattern of iron filings near a wire (Fig. 33-9).

At any point, the direction of $\vec{\mathbf{B}}$ is tangent to the field line at that point. The field is large where the field lines are close together (such as near the wire) and small where the field lines are farther apart (far from the wire). Contrary to electric field lines due to charges, which begin on positive charges and end on negative charges, the magnetic field lines due to currents form continuous loops with no beginning or end. To find the direction of the field lines, we use the right-hand rule: if you were to grasp the wire in your right hand with your thumb in the direction of the current, your fingers would curl around the wire in the direction of the magnetic field.

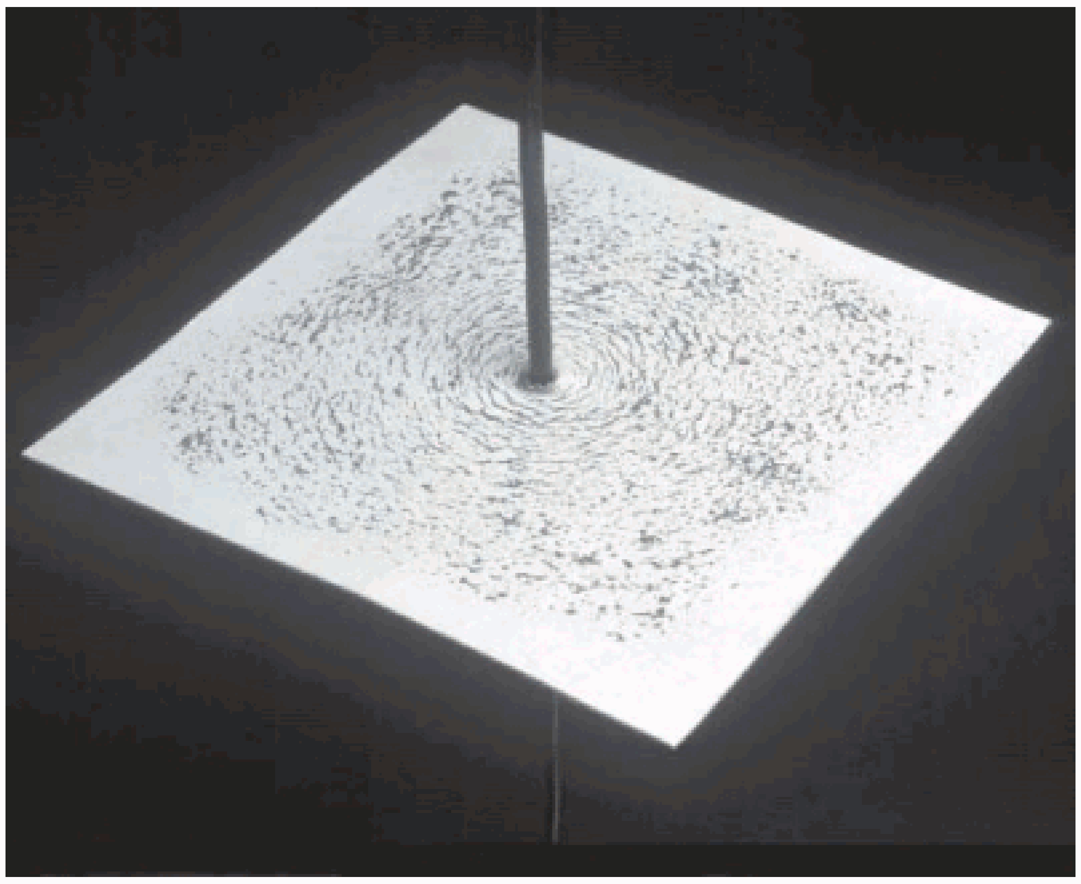


FIGURE 33-9. The vertical wire carries a current, which sets up a magnetic field. Iron filings sprinkled onto cardboard show the pattern of concentric circles that represents the field.

A Circular Current Loop

Figure 33-10 shows a circular loop of radius R carrying a current i. Let us calculate $\vec{\mathbf{B}}$ at a point P on the axis a distance z from the center of the loop.

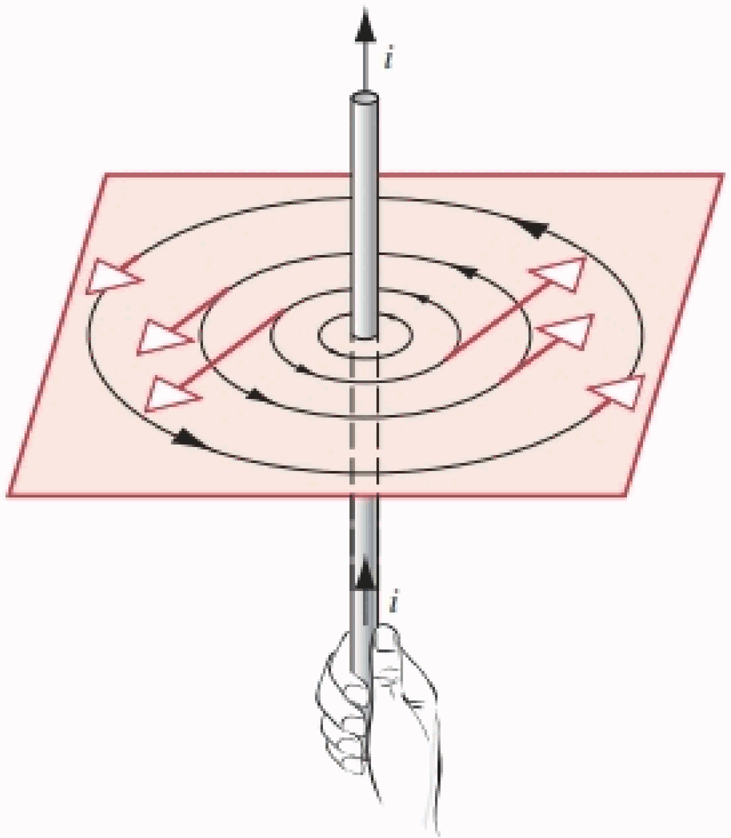


FIGURE 33-8. The lines of the magnetic field are concentric circles for a long, straight, current-carrying wire. Their direction is given by the right-hand rule.

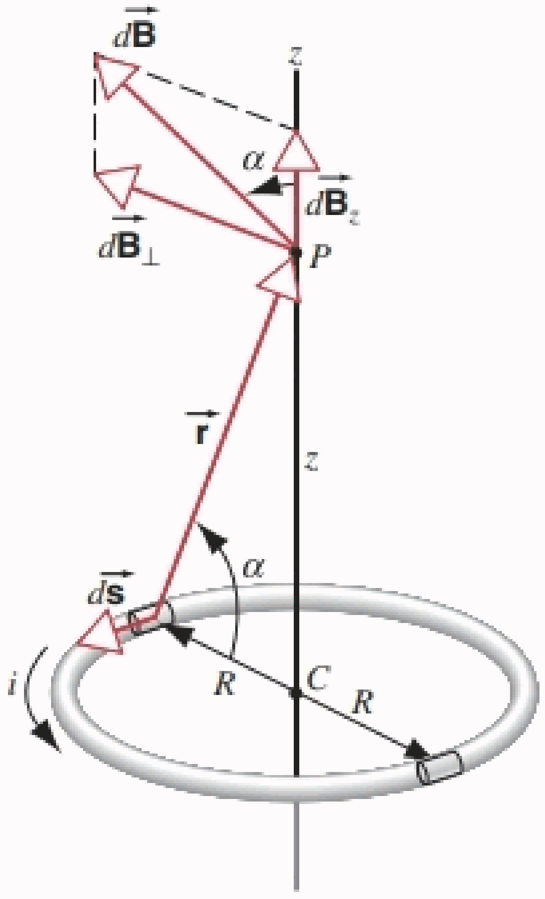


FIGURE 33-10. A circular loop of current. The element $i\,d\vec{\mathbf{s}}$ of the loop sets up a field $d\vec{\mathbf{B}}$ at a point P on the axis of the loop.

The angle ϕ between the current element $i\,d\vec{\mathbf{s}}$ and $\vec{\mathbf{r}}$ is 90°. From the Biot–Savart law, we know that the vector $d\vec{\mathbf{B}}$ for this element is at right angles to the plane formed by $i\,d\vec{\mathbf{s}}$ and $\vec{\mathbf{r}}$ and thus lies at right angles to $\vec{\mathbf{r}}$, as the figure shows.

Let us resolve $d\vec{\mathbf{B}}$ into two components; one, $d\vec{\mathbf{B}}_z$, along the axis of the loop and another, $d\vec{\mathbf{B}}_\perp$, at right angles to the axis. Only $d\vec{\mathbf{B}}_z$ contributes to the total magnetic field $\vec{\mathbf{B}}$ at point P. This follows because the components $d\vec{\mathbf{B}}_z$ for all current elements lie on the axis and add directly; however, the components $d\vec{\mathbf{B}}_\perp$ point in different directions perpendicular to the axis, and the sum of all $d\vec{\mathbf{B}}_\perp$ for the complete loop is zero, from symmetry. (A diametrically opposite current element, indicated in Fig. 33-10, produces the same $d\vec{\mathbf{B}}_z$ but $d\vec{\mathbf{B}}_\perp$ in the opposite direction.) We can therefore replace the vector integral over all $d\vec{\mathbf{B}}$ with an integral over the z components only, and the magnitude of the field is given by

$$B = \int dB_z. \tag{33-14}$$

For the current element in Fig. 33-10, the Biot–Savart law (Eq. 33-9) gives

$$dB = \frac{\mu_0 i}{4\pi}\frac{ds \sin 90^\circ}{r^2}. \tag{33-15}$$

We also have

$$dB_z = dB \cos\alpha,$$

which, combined with Eq. 33-15, gives

$$dB_z = \frac{\mu_0 i \cos\alpha\, ds}{4\pi r^2}. \tag{33-16}$$

Figure 33-10 shows that r and α are not independent of each other. Let us express each in terms of z, the distance from the center of the loop to the point P. The relationships are

$$r = \sqrt{R^2 + z^2}$$

and

$$\cos\alpha = \frac{R}{r} = \frac{R}{\sqrt{R^2 + z^2}}.$$

Substituting these values into Eq. 33-16 for dB_z gives

$$dB_z = \frac{\mu_0 i R}{4\pi(R^2 + z^2)^{3/2}}\, ds. \tag{33-17}$$

Note that i, R, and z have the same values for all current elements. Integrating this equation, we obtain

$$B = \int dB_z = \frac{\mu_0 i R}{4\pi(R^2 + z^2)^{3/2}} \int ds \tag{33-18}$$

or, noting that $\int ds$ is simply the circumference of the loop $(= 2\pi R)$,

$$B = \frac{\mu_0 i R^2}{2(R^2 + z^2)^{3/2}}. \tag{33-19}$$

We can repeat the above calculation to find the field at the center of the loop. In this case $r = R$ everywhere, and the Biot–Savart law gives

$$B = \int dB_z = \frac{\mu_0 i}{4\pi R^2}\int ds. \tag{33-20}$$

If we integrate around the circle, the integral is again $2\pi R$, and so

$$B = \frac{\mu_0 i}{2R}, \tag{33-21}$$

which we could have obtained by setting $z = 0$ in Eq. 33-19. However, we can use this method to obtain a more general result when the current flows not in a complete circle but in an arc of a circle. Suppose the arc subtends an angle θ at its center. Then in Eq. 33-20, the integral gives not the full circumference of the circle but only the arc length $R\theta$ (which is equal to $2\pi R$ when $\theta = 2\pi$). The field at the center of the arc is then

$$B = \frac{\mu_0 i \theta}{4\pi R}. \tag{33-22}$$

The angle θ must be expressed in radians in this equation. The right-hand rule again gives the direction of the magnetic field, which is along the z axis.

If $z \gg R$, so that points close to the loop are not considered, Eq. 33-19 becomes

$$B = \frac{\mu_0 i R^2}{2z^3}. \tag{33-23}$$

This dependence of the field on the inverse cube of the distance reminds us of the electric field of an electric dipole (see Eq. 26-12 and also see Problem 1 of Chapter 26 for the field on the dipole axis). It is often convenient to consider a loop of wire to be a *magnetic dipole.* Just as the electric behavior of many molecules can be characterized in terms of their electric dipole moment, so also the magnetic behavior of atoms can be described in terms of their *magnetic dipole moment.* In the case of atoms, the current loop is due to the circulation of electrons about the nucleus. We discuss the magnetic dipole moments of atoms in Chapter 35.

SAMPLE PROBLEM 33-2. In the Bohr model of the hydrogen atom, the electron circulates around the nucleus in a circular path of radius 5.29×10^{-11} m at a frequency f of 6.60×10^{15} Hz (or rev/s). What value of B is set up at the center of the orbit?

Solution The current is the rate at which charge passes any point on the orbit and is given by

$$i = ef = (1.60 \times 10^{-19}\,\mathrm{C})(6.60 \times 10^{15}\,\mathrm{Hz}) = 1.06 \times 10^{-3}\,\mathrm{A}.$$

The magnetic field B at the center of the orbit is given by Eq. 33-21,

$$B = \frac{\mu_0 i}{2R} = \frac{(4\pi \times 10^{-7}\,\mathrm{T\cdot m/A})(1.06 \times 10^{-3}\,\mathrm{A})}{2(5.29 \times 10^{-11}\,\mathrm{m})} = 12.6\,\mathrm{T}.$$

33-3 TWO PARALLEL CURRENTS

In this section we use long wires carrying parallel (or antiparallel) currents to illustrate two properties of magnetic fields: the addition of the fields due to different wires and the force exerted by one wire on another.

First we consider the vector addition of the fields due to two different parallel wires, as shown in Fig. 33-11. Two wires are perpendicular to the plane of the figure, and they carry currents in opposite directions. We wish to find the magnetic field at point P due to the two wires. The magnetic field lines due to wire 1 form concentric circles about that wire, and the magnitude of the field at the distance r_1 is given by Eq. 33-13, $B = \mu_0 i_1/2\pi r_1$. The direction of $\vec{\mathbf{B}}_1$ is tangent to the circular arc passing through P; equivalently, $\vec{\mathbf{B}}_1$ is perpendicular to $\vec{\mathbf{r}}_1$, the radial vector from the wire to P.

Similarly, the field due to wire 2 is shown in the figure as $\vec{\mathbf{B}}_2$ and is tangent to the circular magnetic field lines and perpendicular to $\vec{\mathbf{r}}_2$. To find the net field at P, we take the vector sum of the fields due to the two individual wires: $\vec{\mathbf{B}} = \vec{\mathbf{B}}_1 + \vec{\mathbf{B}}_2$. The magnitude and direction of the total field can be found using the usual rules for vector addition.

The situation shown in Fig. 33-11 is similar to the method for calculating the total electric field due to two point charges q_1 and q_2: we find the individual fields at point P due to each charge, and then the vector sum gives the total field, $\vec{\mathbf{E}} = \vec{\mathbf{E}}_1 + \vec{\mathbf{E}}_2$. To observe this total electric field at P, we could measure the force on a third charged particle placed at that point. Similarly, to observe the total magnetic field at P in Fig. 33-11, we could measure the force on a charged particle moving through that point or on a third wire carrying a current through P.

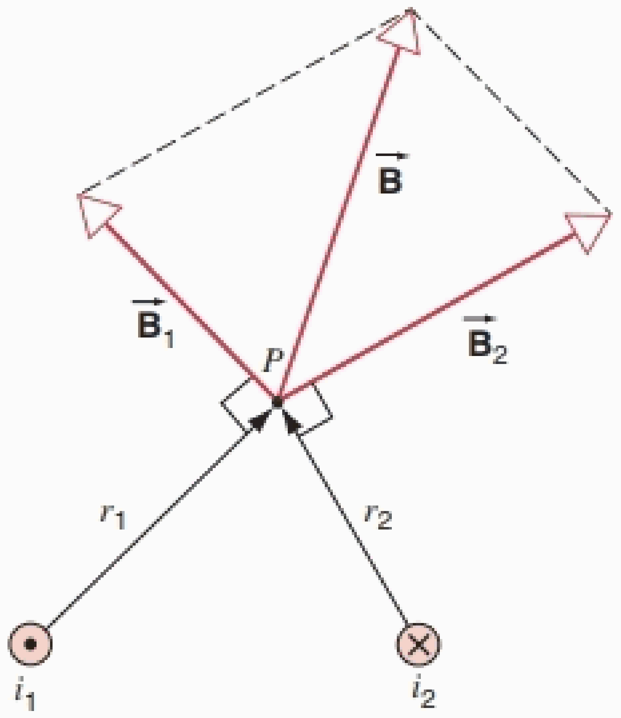


FIGURE 33-11. Two wires carry currents perpendicular to the page; i_1 is out of the page (represented by ⊙, suggesting the pointed tip of an arrow) and i_2 is into the page (represented by ⊗, suggesting the "tailfeathers" of an arrow). The total field at point P is the vector sum of $\vec{\mathbf{B}}_1$ and $\vec{\mathbf{B}}_2$.

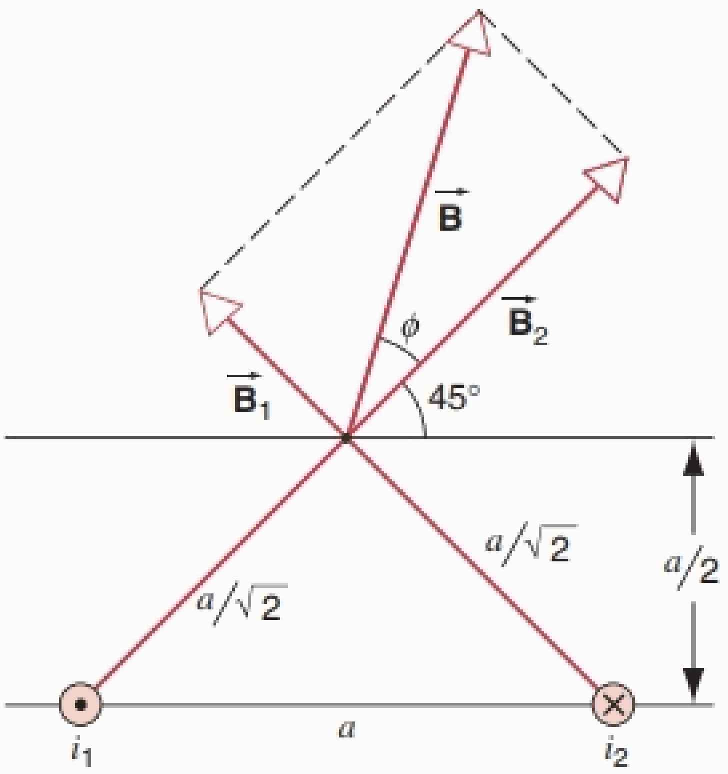


FIGURE 33-12. Sample Problem 33-3. Current i_1 is out of the page and i_2 is into the page.

SAMPLE PROBLEM 33-3. In Fig. 33-12, let $i_1 = 15$ A and $i_2 = 32$ A. The two wires are separated by a distance $a = 5.3$ cm. Find the total magnetic field at a point a distance $a/2$ along a line perpendicular to the line connecting the two wires.

Solution Figure 33-12 shows the geometry and the fields $\vec{\mathbf{B}}_1$ and $\vec{\mathbf{B}}_2$. With $d_1 = d_2 = a/\sqrt{2}$, the magnitudes of the fields are

$$B_1 = \frac{\mu_0 i_1}{2\pi d_1} = \frac{(4\pi \times 10^{-7}\ \mathrm{T\cdot m/A})(15\ \mathrm{A})}{2\pi(0.053\ \mathrm{m})/\sqrt{2}} = 80\ \mu\mathrm{T},$$

$$B_2 = \frac{\mu_0 i_2}{2\pi d_2} = \frac{(4\pi \times 10^{-7}\ \mathrm{T\cdot m/A})(32\ \mathrm{A})}{2\pi(0.053\ \mathrm{m})/\sqrt{2}} = 171\ \mu\mathrm{T}.$$

In the special geometry of Fig. 33-12 the two fields are perpendicular, so

$$B = \sqrt{B_1^2 + B_2^2} = 190\ \mu\mathrm{T}.$$

The angle ϕ between $\vec{\mathbf{B}}$ and $\vec{\mathbf{B}}_2$ is

$$\phi = \tan^{-1}\frac{B_1}{B_2} = 25°,$$

so the angle between $\vec{\mathbf{B}}$ and the horizontal axis is 25° + 45° = 70°.

SAMPLE PROBLEM 33-4. Two long, parallel wires a distance $2b$ apart carry equal currents i in opposite directions, as shown in Fig. 33-13a. Derive an expression for the magnetic field B at a point P on the line connecting the wires and a distance x from the point midway between them.

Solution Study of Fig. 33-13a shows that $\vec{\mathbf{B}}_1$ due to the current i_1 and $\vec{\mathbf{B}}_2$ due to the current i_2 point in the same direction at P. Each is given by Eq. 33-13 ($B = \mu_0 i/2\pi d$) so that

$$B = B_1 + B_2 = \frac{\mu_0 i}{2\pi(b + x)} + \frac{\mu_0 i}{2\pi(b - x)} = \frac{\mu_0 i b}{\pi(b^2 - x^2)}.$$

Inspection of this result shows that (1) B is symmetrical about $x = 0$, (2) B has its minimum value ($= \mu_0 i/\pi b$) at $x = 0$, and (3) $B \rightarrow \infty$ as $x \rightarrow \pm b$. This last conclusion is not correct, because Eq. 33-13 cannot be applied to points inside the wires. In re-

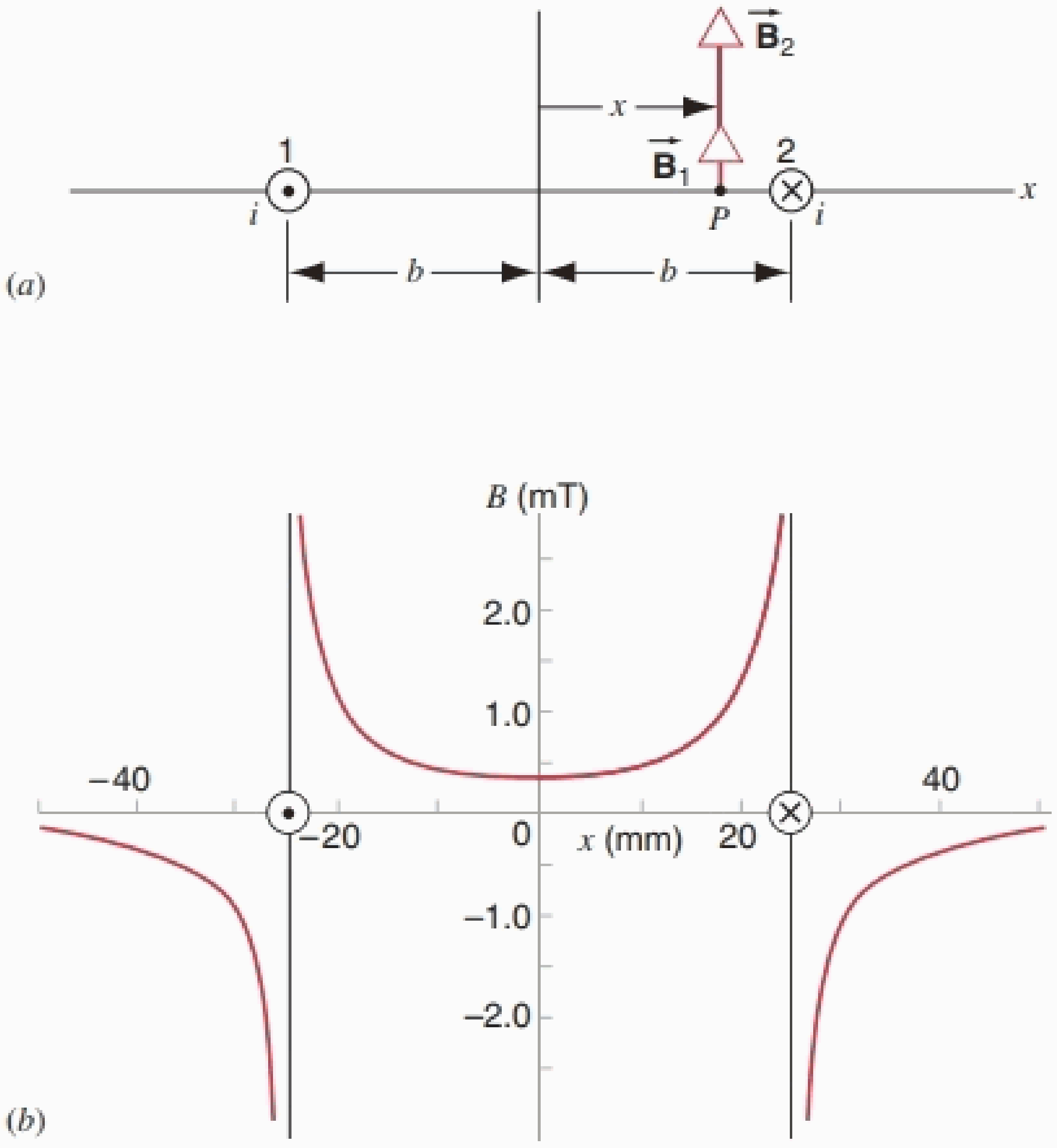


FIGURE 33-13. Sample Problem 33-4. (*a*) The magnetic fields at point *P* due to the currents in wires 1 and 2. (*b*) The resultant field at *P*, calculated for $i = 25$ A and $b = 25$ mm.

ality the field due to each wire would vanish at the center of that wire.

You should show that our result for the combined field remains valid at points where $|x| > b$. Figure 33-13*b* shows the variation of *B* with *x* for $i = 25\ A$ and $b = 25$ mm.

SAMPLE PROBLEM 33-5. Figure 33-14 shows a flat strip of copper of width *a* and negligible thickness carrying a current *i*. Find the magnetic field $\vec{B}$ at point *P*, at a distance *R* from the center of the strip along its perpendicular bisector.

Solution Let us subdivide the strip into long, infinitesimal filaments of width *dx*, each of which may be treated as a wire carrying a current element *di* given by *i*(*dx*/*a*). For the current element in the left half of the strip in Fig. 33-14, the magnitude *dB* of the field at *P* is given by the differential form of Eq. 33-13, or

$$dB = \frac{\mu_0}{2\pi}\frac{di}{d} = \frac{\mu_0}{2\pi}\frac{i(dx/a)}{R\sec\theta},$$

in which $d = R/\cos\theta = R\sec\theta$. Note that the vector $d\vec{B}$ is at right angles to the line marked *d*.

Only the horizontal component of $d\vec{B}$—namely, $dB\cos\theta$—is effective; the vertical component is canceled by the contribution of a symmetrically located current element on the other side of the strip (the second shaded element in Fig. 33-14). Thus *B* at point *P* is given by the (scalar) integral

$$B = \int dB\cos\theta = \int \frac{\mu_0 i(dx/a)}{2\pi R\sec\theta}\cos\theta$$

$$= \frac{\mu_0 i}{2\pi aR}\int \frac{dx}{\sec^2\theta}.$$

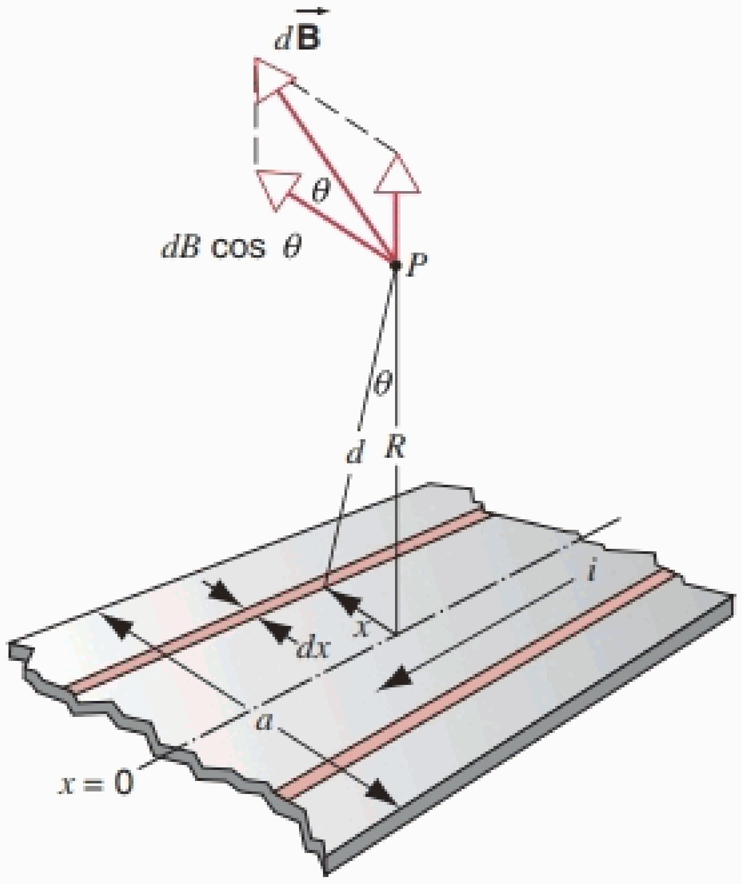


FIGURE 33-14. Sample Problem 33-5. A flat strip of width *a* carries a current *i*.

The variables *x* and θ are not independent, being related by

$$x = R\tan\theta$$

or

$$dx = R\sec^2\theta\, d\theta.$$

The limits on θ are $\pm\alpha$, where $\alpha = \tan^{-1}(a/2R)$. Substituting for *dx* in the expression for *B*, we find

$$B = \frac{\mu_0 i}{2\pi aR}\int \frac{R\sec^2\theta\, d\theta}{\sec^2\theta}$$

$$= \frac{\mu_0 i}{2\pi a}\int_{-\alpha}^{+\alpha} d\theta = \frac{\mu_0 i}{\pi a}\alpha = \frac{\mu_0 i}{\pi a}\tan^{-1}\frac{a}{2R}. \qquad (33\text{-}24)$$

This is the general result for the magnetic field due to the strip.

At points far from the strip, α is a small angle, for which $\alpha \approx \tan\alpha = a/2R$. Thus we have, as an approximate result,

$$B \approx \frac{\mu_0 i}{\pi a}\left(\frac{a}{2R}\right) = \frac{\mu_0}{2\pi}\frac{i}{R}.$$

This result is expected because at distant points the strip cannot be distinguished from a thin wire (see Eq. 33-13).

The Interaction between Parallel Currents

We now consider a different calculation involving two long, straight wires carrying parallel (or antiparallel) currents. As a result of the magnetic field due to one wire at the location of the other wire, a magnetic force is exerted on the second wire. Similarly, the second wire sets up a magnetic field at the location of the first wire that exerts a force on that wire.

In Fig. 33-15, wire 1 carrying current i_1 produces a magnetic field $\vec{B}_1$ whose magnitude at the location of the second wire is, according to Eq. 33-13,

$$B_1 = \frac{\mu_0 i_1}{2\pi d}.$$

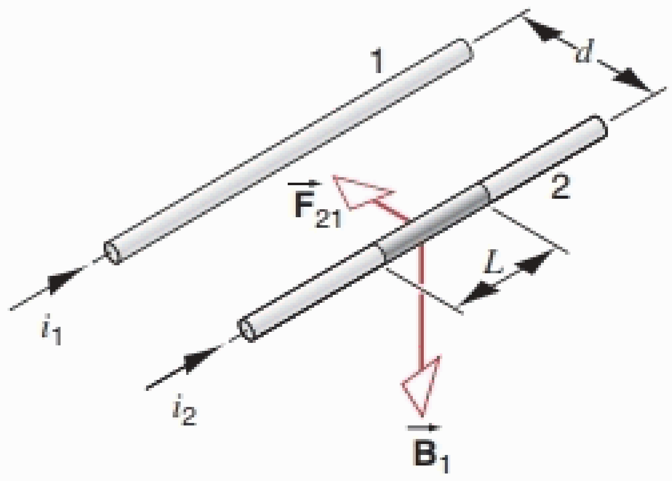


FIGURE 33-15. Two parallel wires carrying currents in the same direction attract each other. The field $\vec{\mathbf{B}}_1$ at wire 2 is that due to the current in wire 1.

The right-hand rule shows that the direction of $\vec{\mathbf{B}}_1$ at wire 2 is down, as shown in the figure.

Wire 2, which carries a current i_2, can thus be considered to be immersed in an *external* magnetic field $\vec{\mathbf{B}}_1$. A length L of this wire experiences a sideways magnetic force $\vec{\mathbf{F}}_{21} = i_2\vec{\mathbf{L}} \times \vec{\mathbf{B}}_1$ of magnitude

$$F_{21} = i_2LB_1 = \frac{\mu_0 L i_1 i_2}{2\pi d}. \qquad (33\text{-}25)$$

The vector rule for the cross product shows that $\vec{\mathbf{F}}_{21}$ lies in the plane of the wires and points toward wire 1 in Fig. 33-15.

We could equally well have started with wire 2 by first computing the magnetic field $\vec{\mathbf{B}}_2$ produced by wire 2 at the site of wire 1 and then finding the force $\vec{\mathbf{F}}_{12}$ exerted on a length L of wire 1 by the field of wire 2. This force on wire 1 would, for parallel currents, point toward wire 2 in Fig. 33-15. The forces that the two wires exert on each other are equal in magnitude and opposite in direction; they form an action–reaction pair according to Newton's third law.

If the currents in Fig. 33-15 were antiparallel, we would find that the forces on the wires would have the opposite directions: the wires would repel one another. The general rule is:

> *Parallel currents attract, and antiparallel currents repel.*

This rule is in a sense opposite to the rule for electric charges, in that like (parallel) currents attract, but like (same sign) charges repel.

The force between long, parallel wires is used to define the ampere. Given two long, parallel wires of negligible circular cross section separated in vacuum by a distance of 1 meter, the ampere is defined as the current in each wire that would produce a force of 2×10^{-7} newtons per meter of length.

SAMPLE PROBLEM 33-6. A long, horizontal, rigidly supported wire carries a current i_a of 96 A. Directly above it and parallel to it is a fine wire that carries a current i_b of 23 A and weighs 0.073 N/m. How far above the lower wire should this second wire be strung if we hope to support it by magnetic repulsion?

Solution To provide a repulsion, the two currents must point in opposite directions. For equilibrium, the magnetic force per unit length must equal the weight per unit length and must be oppositely directed. Solving Eq. 33-25 for d yields

$$d = \frac{\mu_0 i_a i_b}{2\pi(F/L)} = \frac{(4\pi \times 10^{-7}\,\mathrm{T\cdot m/A})(96\,\mathrm{A})(23\,\mathrm{A})}{2\pi(0.073\,\mathrm{N/m})}$$

$$= 6.0 \times 10^{-3}\,\mathrm{m} = 6.0\,\mathrm{mm}.$$

We assume that the diameter of the suspended wire is much smaller than the separation of the two wires. This assumption is necessary because in deriving Eq. 33-25 we tacitly assumed that the magnetic field produced by one wire is uniform for all points within the second wire.

Is the equilibrium of the suspended wire stable or unstable against vertical displacements? This can be tested by displacing the wire vertically and examining how the forces on the wire change. Is the equilibrium stable or unstable against horizontal displacements?

Suppose that the fine wire is suspended *below* the rigidly supported wire. How may it be made to "float"? Is the equilibrium stable or unstable against vertical displacements? Against horizontal displacements?

33-4 THE MAGNETIC FIELD OF A SOLENOID

Sample Problem 33-5 suggests one way to obtain a uniform magnetic field (that is, a magnetic field that does not vary in either magnitude or direction). A flat strip of conductor, carrying a uniformly distributed current i, sets up a magnetic field given by Eq. 33-24. At points very close to the strip ($R \to 0$ and $\tan^{-1} a/2R \to \pi/2$), Eq. 33-24 becomes $B = \mu_0 i/2a$, which is independent of the distance R from the strip. This reminds us of the electric field near a flat plate carrying a uniform charge density, which likewise does not vary in magnitude or direction. In analogy with the parallel-plate capacitor for electric fields, we could create a device with two flat plates carrying equal currents in opposite directions, where the fields would reinforce in the region between the plates and cancel outside the plates.

A more practical way to obtain a nearly uniform magnetic field is to use a *solenoid*. As indicated in Fig. 33-16*a*, a solenoid is a helical winding on a cylindrical core. The wires carry a current i and are wound tightly together, so that there are n windings per unit length along the solenoid.

In this section we calculate the field along the central axis of the solenoid using our previous result for the magnetic field of a circular loop of wire. The calculation of the field off the axis is difficult using the Biot–Savart law, but in the next section we discuss a different and much easier way to calculate the field off axis.

Figure 33-16*b* shows the geometry for calculating the field on the axis. We take the symmetry axis of the solenoid to be the z axis, with the origin at the center of the solenoid. We wish to find the field at point P, which is a distance d

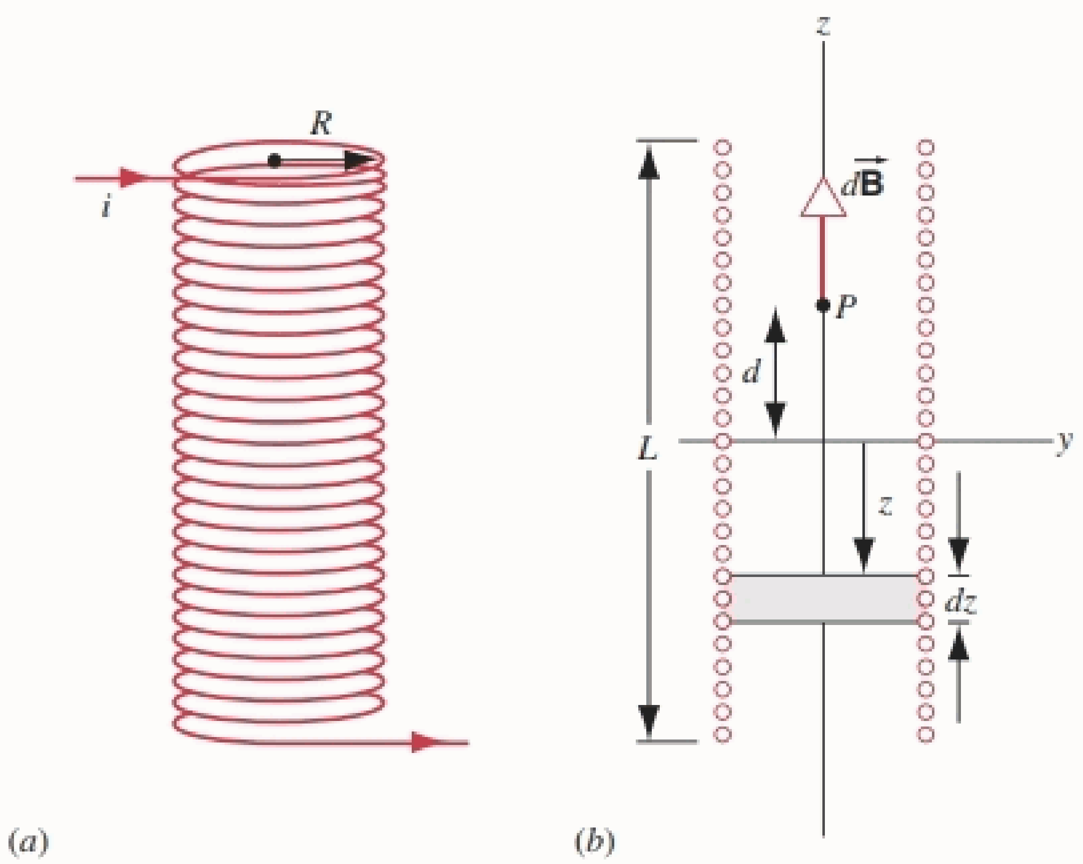


FIGURE 33-16. (*a*) A solenoid. (*b*) A thin ring of width dz gives a field $d\vec{\mathbf{B}}$ at a point P on the z axis.

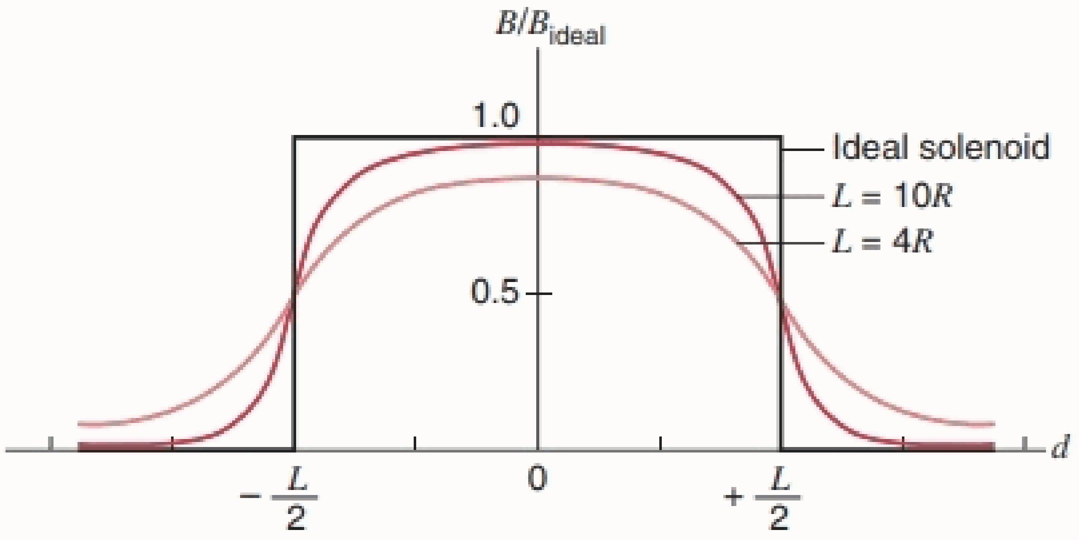


FIGURE 33-17. The magnetic fields of an ideal solenoid and two nonideal solenoids as functions of the distance d from the center.

from the origin along the z axis. We assume that the windings are so narrow that each can be considered as approximately a circular loop of wire, which we assume to be parallel to the xy plane. The solenoid has N turns of wire in a length L, so the number of turns per unit length is $n = N/L$.

Consider a thin ring of width dz. The number of turns in that ring is $n\,dz$, and so the total current carried by the ring is $ni\,dz$, since each turn carries current i. The field at P due to this ring is, using Eq. 33-19,

$$dB = \frac{\mu_0(ni\,dz)R^2}{2[R^2 + (z - d)^2]^{3/2}}, \qquad (33\text{-}26)$$

where $z - d$ is the position of the ring relative to point P. To find the total field due to all such rings, we integrate this expression from $z = -L/2$ to $z = +L/2$. Evaluating the integral (using integral 18 of Appendix I), we obtain

$$\begin{aligned} B &= \frac{\mu_0 niR^2}{2}\int_{-L/2}^{+L/2}\frac{dz}{[R^2 + (z - d)^2]^{3/2}} \\ &= \frac{\mu_0 ni}{2}\left(\frac{L/2 + d}{\sqrt{R^2 + (L/2 + d)^2}} + \frac{L/2 - d}{\sqrt{R^2 + (L/2 - d)^2}}\right). \end{aligned} \qquad (33\text{-}27)$$

This expression gives us the field on the axis of the solenoid at a distance d from its center. It is valid for points inside as well as outside the solenoid. The direction of the field is determined as usual using the right-hand rule, so that if the current is circulating counterclockwise as viewed from above, the field is in the positive z direction.

In an ideal solenoid, the length L is much greater than the radius R. In this case Eq. 33-27 becomes

$$B = \mu_0 ni \qquad \text{(ideal solenoid)}. \qquad (33\text{-}28)$$

As we show in the next section, Eq. 33-28 gives the field of an ideal solenoid at all interior points, off axis as well as on axis, and the field is zero at all points outside the interior of the solenoid.

The field calculated from Eq. 33-27 is plotted as a function of location along the axis in Fig. 33-17 for an ideal solenoid and for two different nonideal solenoids. Note that as the solenoid becomes longer and narrower, thus approaching ideal behavior, the field along the axis becomes more nearly constant and drops more rapidly to zero beyond the ends of the solenoid.

We can begin to understand the field in the interior of a solenoid by considering the "stretched-out" solenoid illustrated in Fig. 33-18. Very close to each wire, the magnetic behavior is nearly that of a long, straight wire, with the field lines forming concentric circles around the wire. This field tends to cancel at points between adjacent wires. The figure suggests that the fields from the individual loops of wire combine to form field lines that are roughly parallel to the solenoid axis in its interior. In the limiting case of the ideal solenoid, the field becomes uniform and parallel to the axis.

At exterior points, such as point P in Fig. 33-18, the field due to the upper part of the solenoid turns (marked ⊙, because the current is out of the page) points to the left and tends to cancel the field due to the lower parts of the solenoid

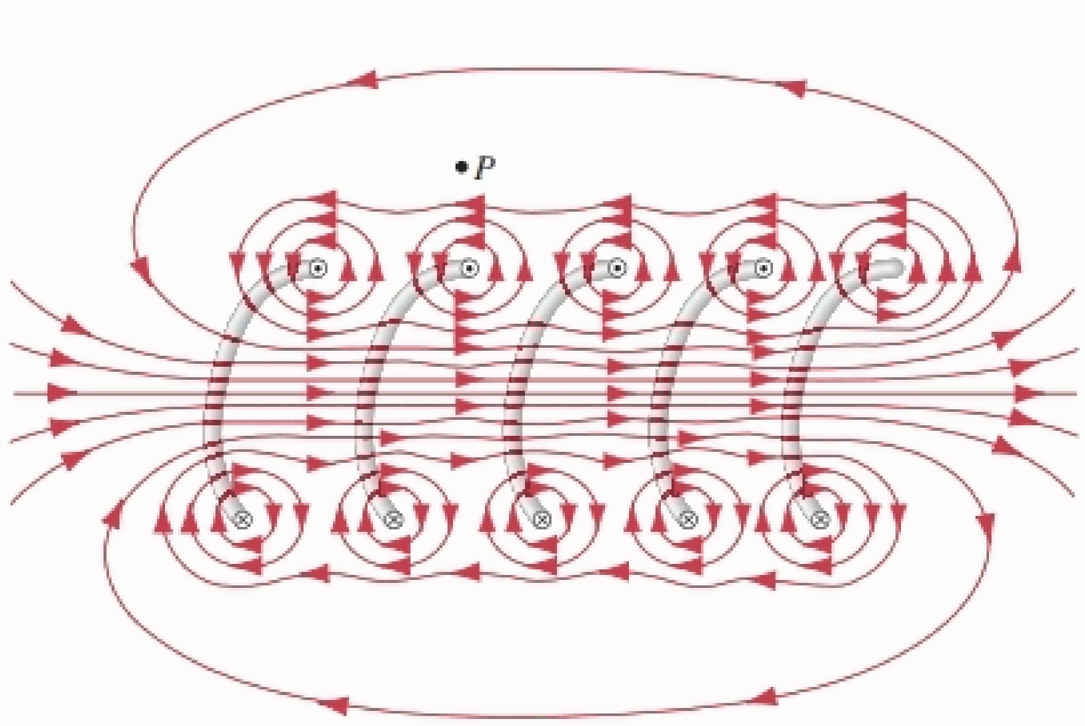


FIGURE 33-18. A section of a solenoid that has been stretched out for this illustration. The magnetic field lines are shown.

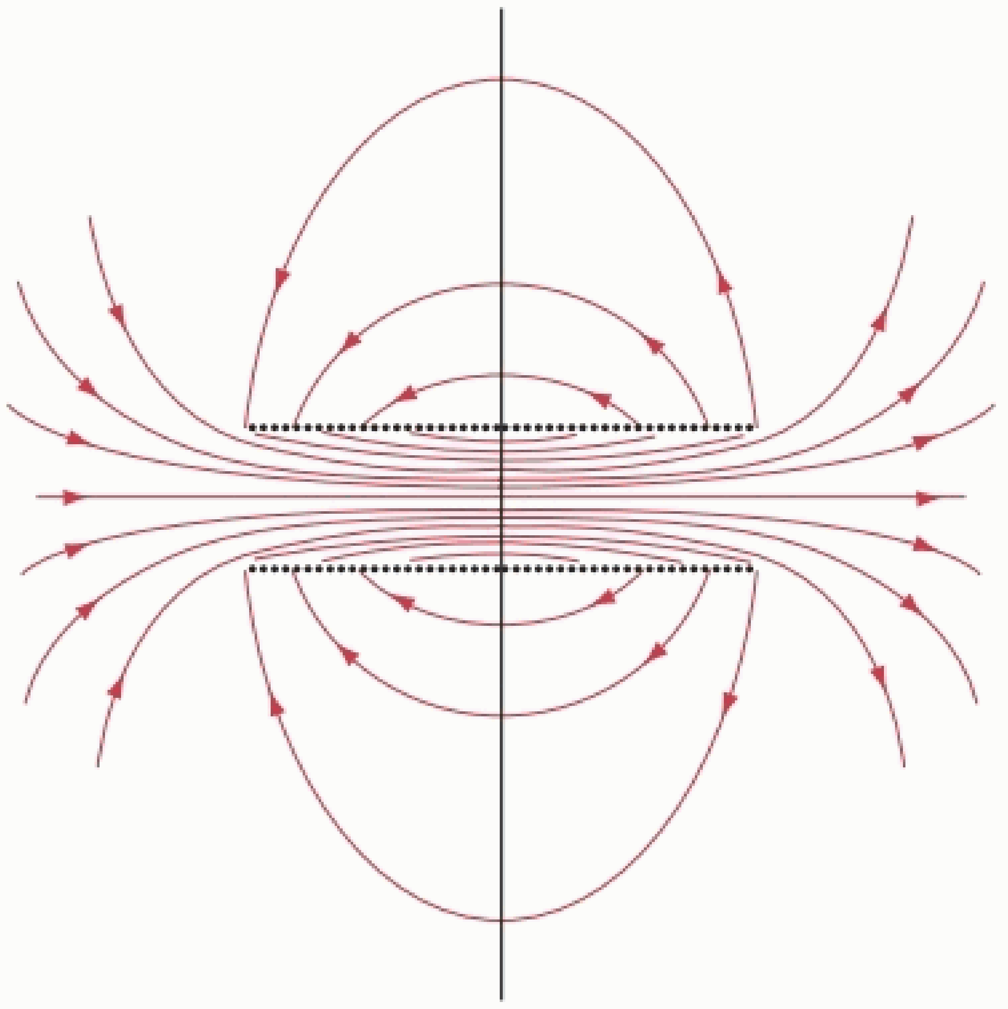


FIGURE 33-19. Magnetic field lines for a solenoid of finite length. Note that the field is stronger (indicated by the greater density of field lines) inside the solenoid than it is outside.

turns (marked ⊗, because the current is into the page), which points to the right near *P*. In the limiting case of the ideal solenoid, the field outside the solenoid is zero. Taking the external field to be zero is a good approximation for a real solenoid if its length is much greater than its radius and if we consider only external points such as *P*. Figure 33-19 shows the magnetic field lines for a nonideal solenoid. You can see from the spacing of the field lines that the field exterior to the solenoid is much weaker than the field in the interior, which is very nearly uniform over the cross section of the solenoid.

The solenoid is for magnetic fields what the parallel-plate capacitor is for electric fields: a relatively simple device capable of producing a field that is approximately uniform. In a parallel-plate capacitor, the electric field is nearly uniform if the plate separation is small compared with the dimensions of the plates, and if we are not too close to the edge of the capacitor. In the solenoid, the magnetic field is nearly uniform if the radius is small compared with the length and if we are not too close to the ends. As shown in Fig. 33-17, even for a length that is only 10 times the radius, the magnetic field is within a few percent of the field of the ideal solenoid over the central half of the device.

SAMPLE PROBLEM 33-7. A solenoid has a length $L = 1.23$ m and an inner diameter $d = 3.55$ cm. It has five layers of windings of 850 turns each and carries a current $i = 5.57$ A. What is B at its center?

Solution With $L/R = 69$, we are safe in regarding this as a nearly ideal solenoid. From Eq. 33-28 we have

$$B = \mu_0 ni = (4\pi \times 10^{-7}\ \mathrm{T \cdot m/A})\left(\frac{5 \times 850\ \text{turns}}{1.23\ \mathrm{m}}\right)(5.57\ \mathrm{A})$$

$$= 2.42 \times 10^{-2}\ \mathrm{T} = 24.2\ \mathrm{mT}.$$

Note that Eq. 33-28 applies even if the solenoid has more than one layer of windings because the diameter of the windings does not enter into the equation.

33-5 AMPÈRE'S LAW

Coulomb's law can be considered a fundamental law of electrostatics; we can use it to calculate the electric field associated with any distribution of electric charges. In Chapter 27, however, we showed that Gauss' law permitted us to solve a certain class of problems, those containing a high degree of symmetry, with ease and elegance. Furthermore, we showed that Gauss' law contained within it Coulomb's law for the electric field of a point charge. We consider Gauss' law to be more basic than Coulomb's law, and Gauss' law is one of the four fundamental (Maxwell) equations of electromagnetism.

The situation in magnetism is similar. Using the Biot–Savart law, we can calculate the magnetic field of any distribution of currents, just as we used Eq. 26-6 or Eqs. 26-13 and 26-14 (which are equivalent to Coulomb's law) to calculate the electric field of any distribution of charges. A more fundamental approach to magnetic fields uses a law that (like Gauss' law for electric fields) takes advantage of the symmetry present in certain problems to simplify the calculation of $\vec{\mathbf{B}}$. This law is considered more fundamental than the Biot–Savart law and leads to another of the four Maxwell equations.

This new result is called *Ampère's law* and is written

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 i. \tag{33-29}$$

You will recall that, in using Gauss' law, we first constructed an imaginary closed surface (a Gaussian surface) that enclosed a certain amount of charge. In using Ampère's law we construct an imaginary closed curve (called an *Ampèrian loop*), as indicated in Fig. 33-20. The left side of Eq. 33-29 tells us to divide the curve into small segments of

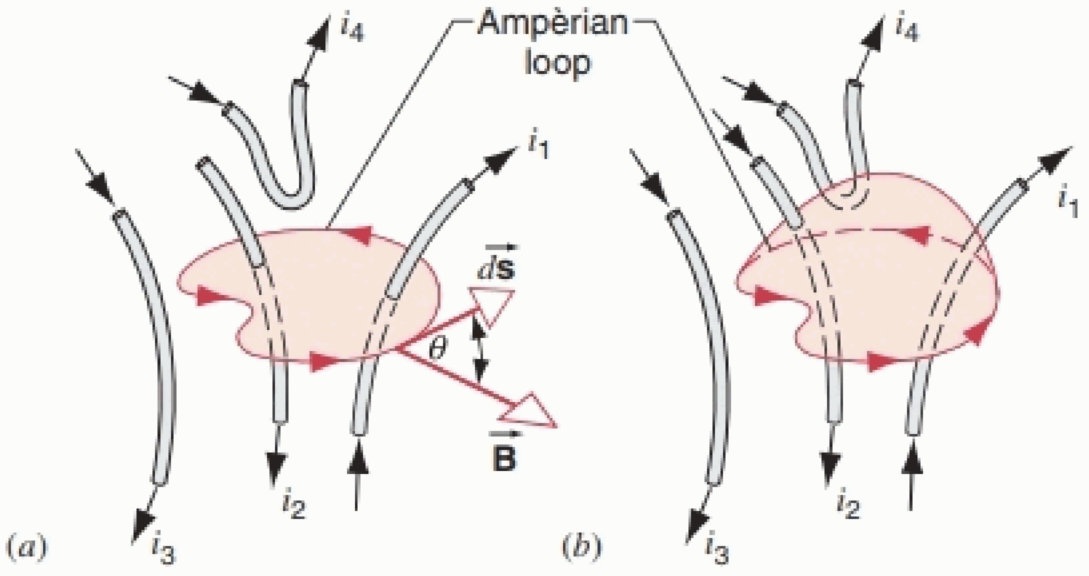


FIGURE 33-20. (*a*) In applying Ampère's law, we integrate around a closed loop. The integral is determined by the net current that passes through the surface bounded by the loop. (*b*) The surface bounded by the loop has been stretched upward.

length $d\vec{s}$. As we travel around the loop (our direction of travel determining the direction of $d\vec{s}$), we evaluate the quantity $\vec{B}\cdot d\vec{s}$ and add (integrate) all such quantities around the loop.

The integral on the left of Eq. 33-29 is called a *line integral.* (Previously we used line integrals in Chapter 11 to calculate work and in Chapter 28 to calculate potential difference.) The circle superimposed on the integral sign reminds us that the line integral is to be evaluated around a *closed* path. Letting θ represent the angle between $d\vec{s}$ and $\vec{B}$, we can write the line integral as

$$\oint \vec{B}\cdot d\vec{s} = \oint B\,ds\cos\theta. \qquad (33\text{-}30)$$

The current i in Eq. 33-29 is the total current "enclosed" by the loop; that is, it is the total current carried by wires that pierce any surface bounded by the loop. In analogy with charges in the case of Gauss' law, currents outside the loop are not included. Figure 33-20*a* shows four wires carrying current. The magnetic field $\vec{B}$ at any point is the net effect of the currents in all wires. However, in the evaluation of the right side of Eq. 33-29, we include only the currents i_1 and i_2, because the wires carrying i_3 and i_4 do not pass through the surface enclosed by the loop. The two wires that pass through the loop carry currents in opposite direction. A right-hand rule is used to assign signs to currents: with the fingers of your right hand in the direction in which the loop is traveled, currents in the direction of your thumb (such as i_1) are taken to be positive, whereas currents in the opposite direction (such as i_2) are taken to be negative. The net current i in the case of Fig. 33-20*a* is thus $i = i_1 - i_2$.

The magnetic field $\vec{B}$ at points on the loop and within the loop certainly depends on the currents i_3 and i_4; however, the *integral* of $\vec{B}\cdot d\vec{s}$ around the loop does *not* depend on currents such as i_3 and i_4 that do not penetrate the surface enclosed by the loop. This is reasonable, because $\vec{B}\cdot d\vec{s}$ for the field established by i_1 or i_2 always has the same sign as we travel around the loop; however, $\vec{B}\cdot d\vec{s}$ for the fields due to i_3 or i_4 change sign as we travel around the loop, and in fact the positive and negative contributions exactly cancel one another.

Changing the shape of the surface without changing the loop does not change these conclusions. In Fig. 33-20*b* the surface has been "stretched" upward so that now the wire carrying current i_4 penetrates the surface. However, note that it does so twice, once moving downward (which would correspond to a contribution $-i_4$ to the total current through the surface, according to our right-hand rule) and once moving upward (which would contribute $+i_4$ to the total). Thus the total current through the surface does not change; this is as expected, because stretching the surface does not change $\vec{B}$ at locations along the fixed loop, and therefore the line integral on the left side of Ampère's law does not change.

Note that including the arbitrary constant of 4π in the Biot–Savart law reduces the constant that appears in Ampère's law to simply μ_0. (A similar simplification of Gauss' law was obtained by including the constant 4π in Coulomb's law.)

For the situation shown in Fig. 33-20, Ampère's law gives

$$\oint B\,ds\cos\theta = \mu_0(i_1 - i_2). \qquad (33\text{-}31)$$

Equation 33-31 is valid for the magnetic field $\vec{B}$ as it varies in both magnitude and direction around the path of the Ampèrian loop. We cannot solve that equation for B unless we can find a way to remove B from the integral. To do so, we use symmetries in the geometry to choose an Ampèrian loop for which B is constant. We used a similar trick in calculating electric fields using Gauss' law.

The following examples illustrate how Ampère's law can be used to calculate magnetic fields in cases with a high degree of symmetry.

Applications of Ampère's Law

A Long, Straight Wire (external points). We can use Ampère's law to find the magnetic field at a distance d from a long, straight wire, a problem we have already solved using the Biot–Savart law.

As illustrated in Fig. 33-21, we choose as our Ampèrian loop a circle of radius d centered on the wire with its plane perpendicular to the wire. From the symmetry of the problem, $\vec{B}$ can depend only on d (and not, for instance, on the angular coordinate around the circle). By choosing a path that is everywhere the same distance from the wire, we know that B is constant around the path.

We know from Oersted's experiments that $\vec{B}$ has only a tangential component. Thus the angle θ is zero, and the line integral becomes

$$\oint B\,ds\cos\theta = B\oint ds = B(2\pi d). \qquad (33\text{-}32)$$

Note that the integral of ds around the path is simply the length of the path, or $2\pi d$ in the case of the circle. The right side of Ampère's law is simply $\mu_0 i$ (taken as positive,

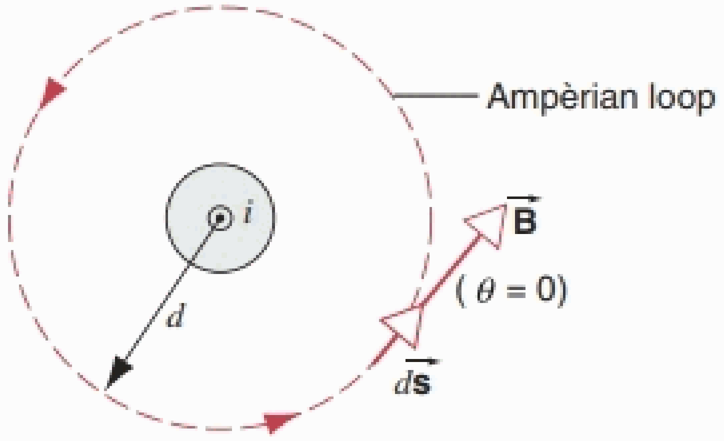


FIGURE 33-21. A circular Ampèrian loop is used to find the magnetic field set up by a current in a long, straight wire. The wire is perpendicular to the page, and the direction of the current is out of the page.

in accordance with the right-hand rule). Ampère's law gives

$$B(2\pi d) = \mu_0 i$$

or

$$B = \frac{\mu_0 i}{2\pi d}.$$

This is identical with Eq. 33-13, a result we obtained (with considerably more effort) using the Biot–Savart law.

A Long, Straight Wire (internal points). We can also use Ampère's law to find the magnetic field *inside* a wire. We assume a cylindrical wire of radius R in which a total current i is distributed uniformly over its cross section. We wish to find the magnetic field at a distance $r < R$ from the center of the wire.

Figure 33-22 shows a circular Ampèrian loop of radius r inside the wire. Symmetry suggests that **B** is constant in magnitude everywhere on the loop and tangent to the loop, so the left side of Ampère's law gives $B(2\pi r)$, exactly as in Eq. 33-32. The right side of Ampère's law involves only the current inside the radius r. If the current is distributed uniformly over the wire, the fraction of the current inside the radius r is the same as the fraction of the area inside r, or $\pi r^2/\pi R^2$. Ampère's law then gives

$$B(2\pi r) = \mu_0 i \frac{\pi r^2}{\pi R^2}, \tag{33-33}$$

where again the right side includes only the fraction of the current that passes through the surface enclosed by the path of integration (the Ampèrian loop).

Solving for B, we obtain

$$B = \frac{\mu_0 i r}{2\pi R^2}. \tag{33-34}$$

At the surface of the wire ($r = R$), Eq. 33-34 reduces to Eq. 33-13 (with $d = R$). That is, both expressions give the same result for the field at the surface of the wire. Figure 33-23 shows how the field depends on r at points both inside and outside the wire.

Equation 33-34 is valid only for the case in which the current is distributed uniformly over the wire. If the current density depends on r, a different result is obtained (see Problem 13). However, Eq. 33-13 for the field outside the wire remains valid whether the current density is constant or a function of r.

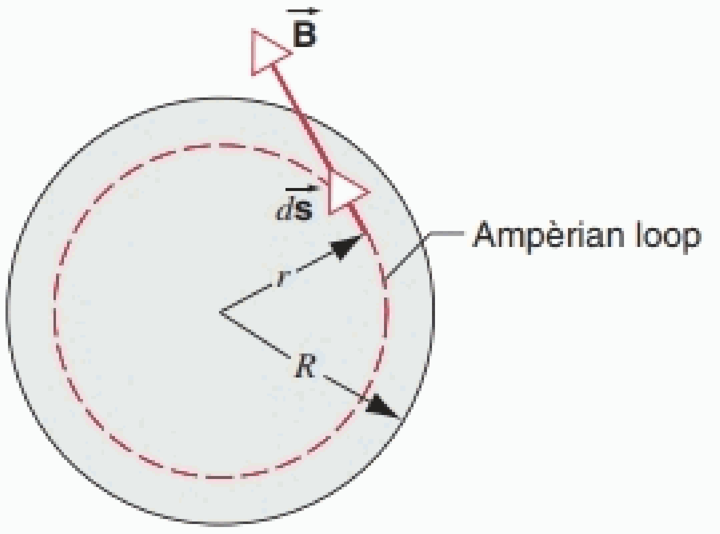


FIGURE 33-22. A long, straight wire carries a current that is emerging from the page and is uniformly distributed over the circular cross section of the wire. A circular Ampèrian loop is drawn inside the wire.

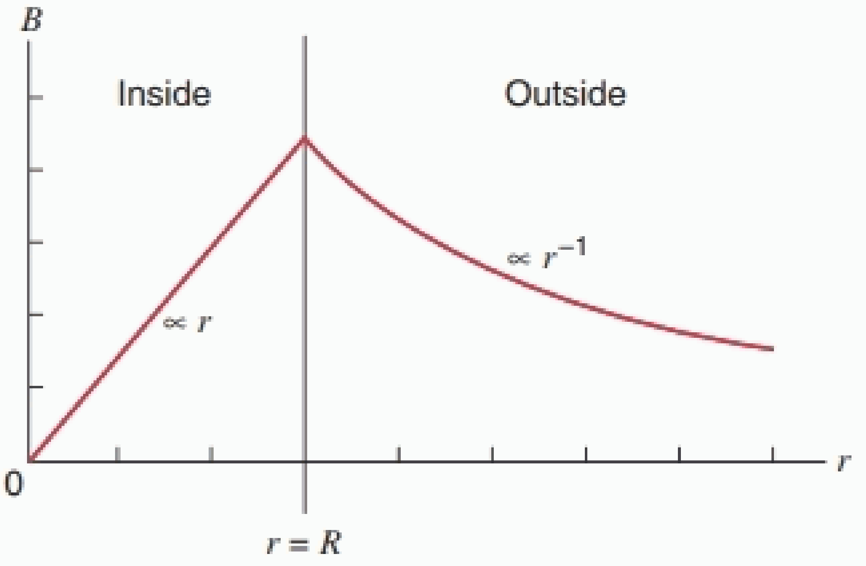


FIGURE 33-23. The magnetic field calculated for the wire shown in Fig. 33-22. Note that the largest field occurs at the surface of the wire.

A Solenoid. We consider an ideal solenoid as shown in Fig. 33-24 and choose an Ampèrian loop in the shape of the rectangle *abcda.* In this analysis we assume that the magnetic field is parallel to the axis of this ideal solenoid and constant in magnitude along line *ab*. As we shall prove, the field is also uniform in the interior (independent of the distance of *ab* from the central axis), as suggested by the equal spacing of the field lines in Fig. 33-24.

The left side of Ampère's law can be written as the sum of four integrals, one for each path segment:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \int_a^b \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} + \int_b^c \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} + \int_c^d \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} + \int_d^a \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}}. \tag{33-35}$$

The first integral on the right is Bh, where B is the magnitude of $\vec{\mathbf{B}}$ inside the solenoid and h is the arbitrary length of the path from a to b. Note that path ab, though parallel to the solenoid axis, need not coincide with it. The second and fourth integrals in Eq. 33-35 are zero because for every element of these paths $\vec{\mathbf{B}}$ is either at right angles to the path (for points inside the solenoid) or is zero (for points outside). In either case, $\vec{\mathbf{B}} \cdot d\vec{\mathbf{s}}$ is zero, and the integrals vanish. The third integral, which includes the part of the rectangle that lies outside the solenoid, is zero because we have taken $\vec{\mathbf{B}}$ as zero for all external points for an ideal solenoid.

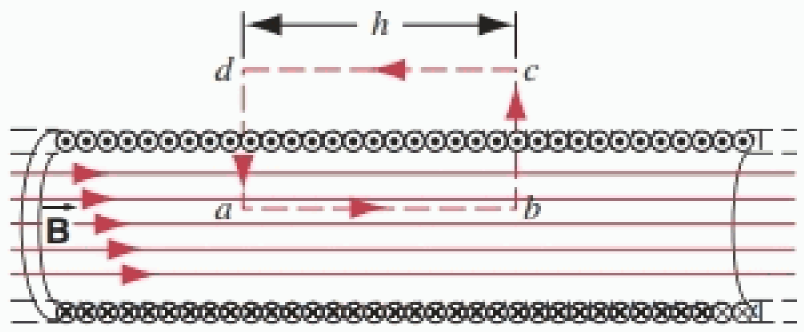


FIGURE 33-24. An Ampèrian loop (the rectangle *abcd*) is used to calculate the magnetic field of this long, idealized solenoid.

For the entire rectangular path, $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}}$ has the value Bh. The net current i that passes through the rectangular Ampèrian loop is not the same as the current in the solenoid because the windings pass through the loop more than once. Let n be the number of turns per unit length; then nh is the number of turns inside the loop, and the total current, passing through the rectangular Ampèrian loop of Fig. 33-24 is nhi. Ampère's law then becomes

$$Bh = \mu_0 nhi$$

or

$$B = \mu_0 ni.$$

This result agrees with Eq. 33-28, which referred only to points on the central axis of the solenoid. Because line ab in Fig. 33-24 can be located at any distance from the axis, we can now conclude that the magnetic field inside an ideal solenoid is uniform over its cross section.

A Toroid. Figure 33-25 shows a toroid, which we may consider to be a solenoid bent into the shape of a doughnut. We can use Ampère's law to find the magnetic field at interior points.

From symmetry, the lines of $\vec{\mathbf{B}}$ form concentric circles inside the toroid, as shown in the figure. Let us choose a concentric circle of radius r as an Ampèrian loop and traverse it in the clockwise direction. Ampère's law yields

$$B(2\pi r) = \mu_0 iN,$$

where i is the current in the toroid windings and N is the total number of turns. This gives

$$B = \frac{\mu_0 iN}{2\pi r}. \tag{33-36}$$

In contrast to the solenoid, B is not constant over the cross section of a toroid. You should be able to show, from Ampère's law, that $B = 0$ for points outside an ideal toroid and in the central cavity.

Close inspection of Eq. 33-36 justifies our earlier statement that a toroid is "a solenoid bent into the shape of a doughnut." The denominator in Eq. 33-36, $2\pi r$, is the central circumference of the toroid, and $N/2\pi r$ is just n, the number of turns per unit length. With this substitution, Eq. 33-36 reduces to $B = \mu_0 in$, the equation for the magnetic field in the central region of a solenoid.

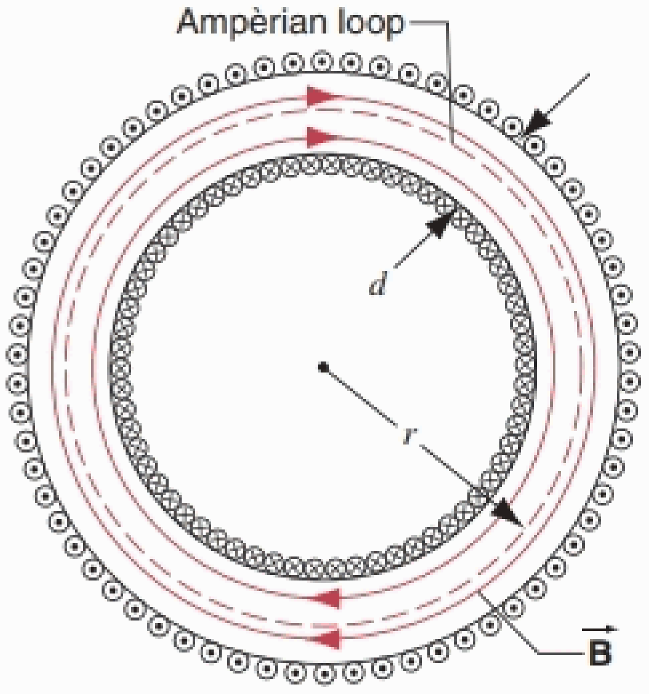


FIGURE 33-25. A toroid. The interior field can be found using the circular Ampèrian loop shown.

The direction of the magnetic field within a toroid (or a solenoid) follows from the right-hand rule: curl the fingers of your right hand in the direction of the current; your extended right thumb then points in the direction of the magnetic field.

Toroids form the central feature of the *tokamak,* a device showing promise as the basis for a fusion power reactor. We discuss its mode of operation in Chapter 51.

The Field Outside a Solenoid (Optional). We have so far neglected the field outside the solenoid, but even for an ideal solenoid, the field at points outside the winding is not zero. Figure 33-26 shows an Ampèrian path in the shape of a circle of radius r. Because the solenoid windings are helical, one turn of the winding pierces the surface enclosed by the circle. The product $\vec{\mathbf{B}} \cdot d\vec{\mathbf{s}}$ for this path depends on the tangential component of the field B_t, and thus Ampère's law gives

$$B_t(2\pi r) = \mu_0 i$$

or

$$B_t = \frac{\mu_0 i}{2\pi r}, \tag{33-37}$$

which is the same field (in magnitude *and* direction) that would be set up by a straight wire. Note that the windings, in addition to carrying current around the surface of the solenoid, also carry current from left to right in Fig. 33-26, and in this respect the solenoid behaves like a straight wire at points outside the windings.

The tangential field is much smaller than the interior field (Eq. 33-28), as we can see by taking the ratio

$$\frac{B_t}{B} = \frac{\mu_0 i/2\pi r}{\mu_0 ni} = \frac{1}{2\pi rn}. \tag{33-38}$$

Suppose the solenoid consists of one layer of turns in which the wires are touching one another, as in Fig. 33-24. Every interval along the solenoid of length equal to the diameter D of the wire contains one turn, and thus the number of turns per unit length n must be $1/D$. The ratio thus becomes

$$\frac{B_t}{B} = \frac{D}{2\pi r}. \tag{33-39}$$

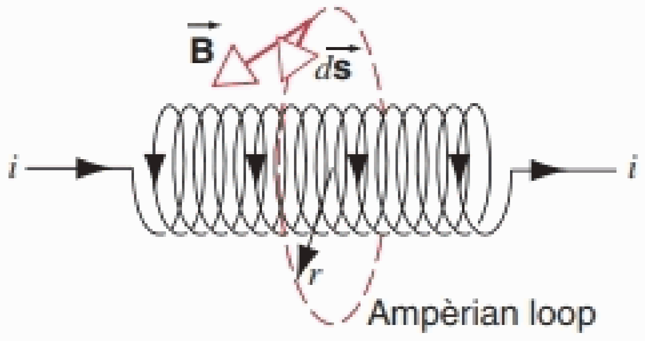


FIGURE 33-26. A circular Ampèrian loop of radius r is used to find the tangential field external to a solenoid.

For a typical wire, $D = 0.1$ mm. The distance r to exterior points must be at least as large as the radius of the solenoid, which might be a few centimeters. Thus $B_t/B \leq 0.001$, and the tangential exterior field is indeed negligible compared with the interior field along the axis. We are therefore safe in neglecting the exterior field.

By drawing an Ampèrian circle similar to that of Fig. 33-26 but with radius smaller than that of the solenoid, you should be able to show that the tangential component of the interior field is zero. ■

33-6 ELECTROMAGNETISM AND FRAMES OF REFERENCE (Optional)

Figure 33-27*a* shows a particle carrying a positive charge q at rest near a long, straight wire that carries a current i. We view the system from a frame of reference S in which the wire is at rest. Inside the wire are negative electrons moving with the drift velocity $\vec{\mathbf{v}}_d$ and positive ion cores at rest. In any given length of wire, the number of electrons equals the number of ion cores, and the net charge is zero. The electrons can instantaneously be considered as a line of negative charge, which sets up an electric field at the location of q according to Eq. 26-17:

$$E = \frac{\lambda_-}{2\pi\epsilon_0 r},$$

where λ_- is the linear charge density of electrons (a negative number). The positive ion cores also set up an electric field given by a similar expression, depending on the linear charge density λ_+ of positive ions. Because the charge densities are of equal magnitude and opposite sign, $\lambda_+ + \lambda_- = 0$, and the net electric field that acts on the particle is zero.

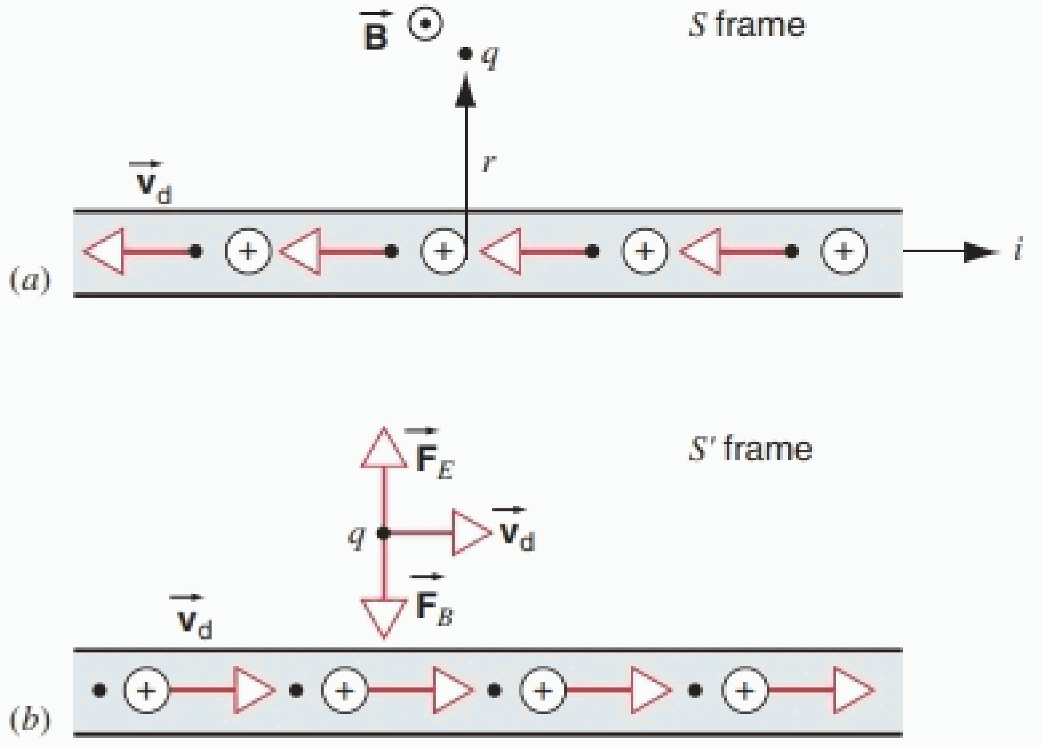


FIGURE 33-27. (*a*) A particle of charge q is at rest in equilibrium near a wire carrying a current i. The situation is viewed from a reference frame S at rest with respect to the particle. (*b*) The same situation viewed from a frame S' that is moving with the drift velocity of the electrons in the wire. The particle is also in equilibrium in this frame under the influence of the two forces $\vec{\mathbf{F}}_E$ and $\vec{\mathbf{F}}_B$.

There is a nonzero magnetic field at the location of the particle, but because the particle is at rest, there is no magnetic force. Therefore no net force of electromagnetic origin acts on the particle in this frame of reference.

Now let us consider the situation from the perspective of a frame of reference S' moving parallel to the wire with velocity $\vec{\mathbf{v}}_d$ (the drift velocity of the electrons). Figure 33-27*b* shows the situation in this frame of reference, in which the electrons are at rest and the ion cores move to the right with velocity $\vec{\mathbf{v}}_d$. Clearly, in this case the particle, being in motion, experiences a magnetic force $\vec{\mathbf{F}}_B$ as shown in the figure.

Observers in different inertial frames must agree that if there is no acceleration of the charge q in the S frame, there must also be no acceleration in the S' frame. The particle must therefore experience no net force in S', and so there must be another force in addition to $\vec{\mathbf{F}}_B$ that acts on the particle to give a net force of zero.

This additional force that acts in the S' frame must be of electric origin. Consider in Fig. 33-27*a* a length L of the wire. We can imagine that length of the wire to consist of two measuring rods, a positively charged rod (the ions) at rest and a negatively charged rod (the electrons) in motion. The two rods have the same length (in S) and contain the same number of charges. When we transform those rods into S', we find that the rod of negative charge has a greater length in S'. In S, this moving rod has its *contracted length,* according to the relativistic effect of length contraction we considered in Section 20-3. In S', it is at rest and has its *proper length,* which is longer than the contracted length in S. The negative linear charge density λ'_- in S' is smaller in magnitude than that in S (that is, $|\lambda'_-| < |\lambda_-|$), because the same amount of charge is spread over a greater length in S'.

For the positive charges, the situation is opposite. In S, the positive charges are at rest, and the rod of positive charge has its proper length. In S', it is in motion and has a shorter, contracted length. The linear density λ'_+ of positive charge in S' is greater than that in S ($\lambda'_+ > \lambda_+$), because the same amount of charge is spread over a shorter length. We therefore have the following relationships for the charge densities:

$$\text{in } S: \qquad \lambda_+ = |\lambda_-|,$$

$$\text{in } S': \qquad \lambda'_+ > |\lambda'_-|.$$

The charge q experiences the electric fields due to a line of positive charge and a line of negative charge. In S', these fields do not cancel, because the linear charge densities are different. The electric field at q in S' is therefore that due to a net linear density of positive charge, and q is repelled from the wire. The electric force $\vec{\mathbf{F}}_E$ on q opposes the magnetic force $\vec{\mathbf{F}}_B$, as shown in Fig. 33-27*b*. A detailed calculation* shows that the resulting electric force is

---

*See, for example, R. Resnick, *Introduction to Special Relativity* (Wiley, 1968), Chapter 4.

exactly equal to the magnetic force, and the net force in S' is zero. Thus the particle experiences no acceleration in either reference frame. We can extend this result to situations other than the special case we considered here, in which S' moves at velocity $\vec{\mathbf{v}}_d$ with respect to S. In other frames of reference, the electric force and the magnetic force have values different from their values in S'; however, in every frame they are equal and opposite to one another and the net force on the particle is zero in *every* frame of reference.

This is a remarkable result. According to special relativity, electric and magnetic fields do not have separate existences. A field that is purely electric or purely magnetic in one frame of reference has both electric and magnetic components in another frame. Using relativistic transformation equations, we can easily pass back and forth from one frame to another, and we can often solve difficult problems by choosing a frame of reference in which the fields have a simpler character and then transforming the result back to the original frame. Special relativity can be of great practical value in solving such problems, because the techniques of special relativity may turn out to be simpler than the classical techniques.

In mathematical language, we say that the laws of electromagnetism (Maxwell's equations) are invariant with respect to the Lorentz transformation. Recall our discussion in Section 11-6 about *invariant* physical laws: we write down the law in one frame of reference, transform to another frame, and obtain a law of exactly the same mathematical form. For example, Gauss' law, one of the four Maxwell equations, has exactly the same form in every frame of reference.

Einstein's words are direct and to the point: "The force acting on a body in motion in a magnetic field is nothing else but an electric field." (In fact, Einstein's original 1905 paper, in which he first presented the ideas of special relativity, was titled "On the Electrodynamics of Moving Bodies.") In this context, we can regard magnetism as a relativistic effect, depending on the velocity of the charge relative to the observer. However, unlike other relativistic effects, it has substantial observable consequences at speeds far smaller than the speed of light. ■

MULTIPLE CHOICE

33-1 The Magnetic Field due to a Moving Charge

1. Two positive charges q_1 and q_2 are moving to the right in Fig. 33-28.

(*a*) What is the direction of the force on charge q_1 due to the magnetic field produced by q_2?
(A) Into the page (B) Out of the page
(C) Up the page (D) Down the page

(*b*) What is the direction of the force on charge q_2 due to the magnetic field produced by q_1?
(A) Into the page (B) Out of the page
(C) Up the page (D) Down the page

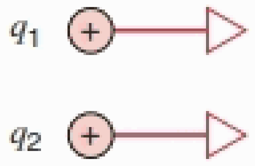


FIGURE 33-28. Multiple-choice question 1.

33-2 The Magnetic Field of a Current

2. Consider the magnitude of the magnetic field $B(z)$ on the axis of a circular current loop.

(*a*) $B(z)$ will be a maximum where
(A) $z = 0$. (B) $0 < |z| < \infty$. (C) $|z| = \infty$.
(D) (A) and (C) are correct.

(*b*) $B(z)$ can be zero where
(A) $z = 0$. (B) $0 < |z| < \infty$. (C) $|z| = \infty$.
(D) (A) and (C) are correct.

3. The negatively charged disk in Fig. 33-29 is rotated clockwise. What is the direction of the magnetic field at point A in the plane of the disk?
(A) Into the page (B) Out of the page
(C) Up the page (D) Down the page

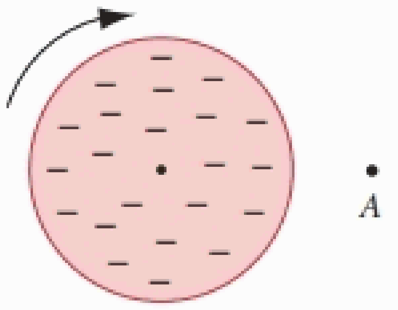


FIGURE 33-29. Multiple-choice question 3.

4. A loop of wire of length L carrying a current i can be wound once as in Fig. 33-30*a*, or twice as in Fig. 33-30*b*. The ratio of the magnetic field strength B_1 at the center of the single loop to the strength B_2 at the center of the double loop is
(A) 2. (B) 1. (C) 1/2. (D) 1/4.

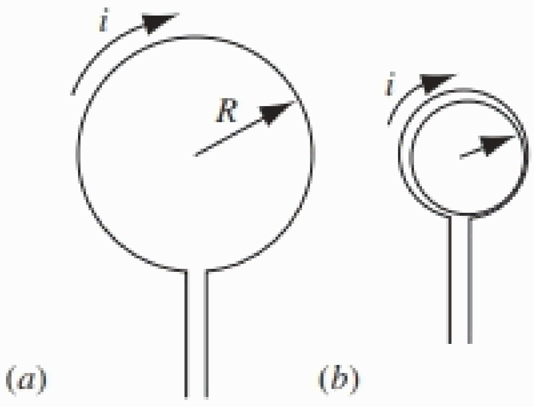


FIGURE 33-30. Multiple-choice question 4.

33-3 Two Parallel Currents

5. A long, straight wire carries a current to the north. A second long, straight wire 0.5 m vertically above the first wire carries an identical current to the east. Both wires are long enough to be considered infinite in length.

(*a*) What is the direction of the net force on the top wire because of the current in the bottom wire?

(A) Up (B) Down (C) North (D) South
(E) The net force is zero.

(*b*) What is the direction of the torque on the top wire because of the current in the bottom wire?

(A) Up (B) Down (C) North (D) South
(E) The torque is zero.

6. Two parallel currents are directed out of the page. Compare the magnitude of the magnetic field B_2 at any arbitrary point equidistant from the wires to the magnitude of the field B_1 at that point from one wire alone.

(A) $B_2 > B_1$ for all equidistant points.
(B) $B_2 = B_1$ for all equidistant points.
(C) $B_2 < B_1$ for all equidistant points.
(D) $B_2 > B_1$ for closer equidistant points only.
(E) $B_2 < B_1$ for closer equidistant points only.

7. Antiparallel currents are directed so that one is out of the page and the other is into the page. Compare the magnitude of the magnetic field B_2 at any arbitrary point equidistant from the wires to the magnitude of the field B_1 at that point from one wire alone.

(A) $B_2 > B_1$ for all equidistant points.
(B) $B_2 = B_1$ for all equidistant points.
(C) $B_2 < B_1$ for all equidistant points.
(D) $B_2 > B_1$ for closer equidistant points only.
(E) $B_2 < B_1$ for closer equidistant points only.

33-4 The Magnetic Field of a Solenoid

8. A metal "Slinky" can be used as a solenoid. The "Slinky" is stretched slightly, and a current is passed through it. Will the resulting magnetic field cause the "Slinky" to collapse or to stretch out further?

(A) Collapse (B) Stretch out further
(C) Neither, the magnetic field is zero outside a solenoid.
(D) The answer depends on the direction of the current.

9. Consider a solenoid with $R \ll L$. The magnetic field at the center of the solenoid is B_0. A second solenoid is constructed that has twice the radius, twice the length, and carries twice the current as the original solenoid, but has the same number of turns per meter. The magnetic field at the center of the second solenoid is

(A) $B_0/2$. (B) B_0. (C) $2B_0$. (D) $4B_0$.

10. How does the magnetic field $B(z)$ behave for points z along the axis of a solenoid with $z \gg L$, where L is the length of the solenoid?

(A) $B(z)$ is constant. (B) $B(z) \propto z^{-1}$
(C) $B(z) \propto z^{-2}$ (D) $B(z) \propto z^{-3}$

33-5 Ampère's Law

11. Solve, without integrating,

$$\int_{-\infty}^{\infty} B\,dz,$$

where B is the magnetic field along the axis of a circular loop of current as given by Eq. 33-19. What is the result?

(A) $i/2R$ (B) $2\mu_0 i$ (C) $\mu_0 i$
(D) The expression cannot be solved without integrating.

12. What is $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}}$ for the path shown in Fig. 33-31?

(A) $-8\pi \times 10^{-7}\,\mathrm{T \cdot m}$ (B) $-4\pi \times 10^{-7}\,\mathrm{T \cdot m}$
(C) $+8\pi \times 10^{-7}\,\mathrm{T \cdot m}$ (D) $+32\pi \times 10^{-7}\,\mathrm{T \cdot m}$

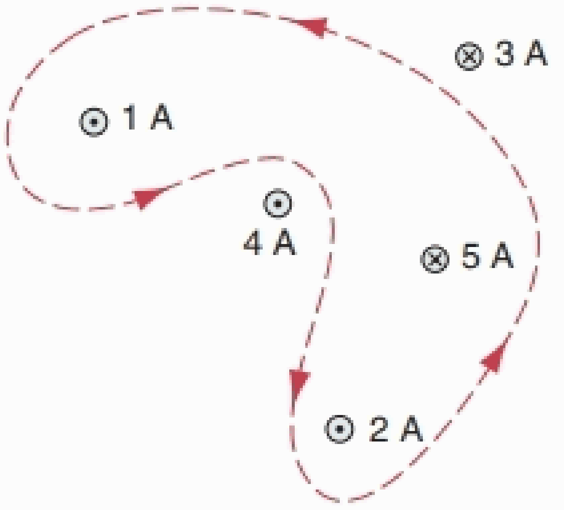


FIGURE 33-31. Multiple-choice question 12.

13. What is $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}}$ for the path shown in Fig. 33-32?

(A) $+56\pi \times 10^{-7}\,\mathrm{T \cdot m}$ (B) $-24\pi \times 10^{-7}\,\mathrm{T \cdot m}$
(C) $+328\pi \times 10^{-7}\,\mathrm{T \cdot m}$ (D) $+80\pi \times 10^{-7}\,\mathrm{T \cdot m}$

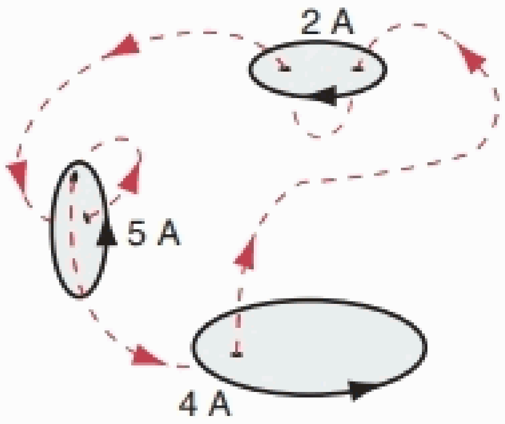


FIGURE 33-32. Multiple-choice question 13.

14. Is $B = \mu_0 n i$ true for infinite solenoids that have noncircular cross sections?

(A) It is a reasonable approximation for cross sections close to circles.
(B) It is a reasonable approximation for any cross-sectional shape.
(C) It is true for cross-sectional shapes of sufficient symmetry (such as equilateral triangles or squares).
(D) It is true for any cross-sectional shape.

33-6 Electromagnetism and Frames of Reference

QUESTIONS

1. A beam of 20-MeV protons emerges from a cyclotron. Do these particles cause a magnetic field?

2. Discuss analogies and differences between Coulomb's law and the Biot–Savart law.

3. Consider a magnetic field line. Is the magnitude of $\vec{\mathbf{B}}$ constant or variable along such a line? Can you give an example of each case?

4. In electronics, wires that carry equal but opposite currents are often twisted together to reduce their magnetic effect at distant points. Why is this effective?

5. Consider two charges, first (*a*) of the same sign and then (*b*) of opposite signs, that are moving along separated parallel paths with the same velocity. Compare the directions of the mutual electric and magnetic forces in each case.

6. Is there any way to set up a magnetic field other than by causing charges to move?

7. Give details of three ways in which you can measure the magnetic field $\vec{\mathbf{B}}$ at a point P, a perpendicular distance r from a long, straight wire carrying a constant current i. Base them on (*a*) projecting a particle of charge q through point P with velocity $\vec{\mathbf{v}}$ parallel to the wire; (*b*) measuring the force per unit length exerted on a second wire, parallel to the first wire and carrying a current i'; (*c*) measuring the torque exerted on a small magnetic dipole located a perpendicular distance r from the wire.

8. Is $\vec{\mathbf{B}}$ uniform for all points within a circular loop of wire carrying a current? Explain.

9. Two long, parallel conductors carry equal currents i in the same direction. Sketch roughly the resultant lines of $\vec{\mathbf{B}}$ due to the action of both currents. Does your figure suggest an attraction between the wires?

10. A current is sent through a vertical spring from whose lower end a weight is hanging. What will happen?

11. Equation 33-13 ($B = \mu_0 i/2\pi d$) suggests that a strong magnetic field is set up at points near a long wire carrying a current. Since there is a current i and magnetic field $\vec{\mathbf{B}}$, why is there not a force on the wire in accord with the equation $\vec{\mathbf{F}}_B = i\vec{\mathbf{L}} \times \vec{\mathbf{B}}$?

12. Two long, straight wires pass near one another at right angles. If the wires are free to move, describe what happens when currents are sent through both of them.

13. Two fixed wires cross each other perpendicularly so that they do not actually touch but are close to each other, as shown in Fig. 33-33. Equal currents i exist in each wire in the directions indicated. In what region(s) will there be some points of zero net magnetic field?

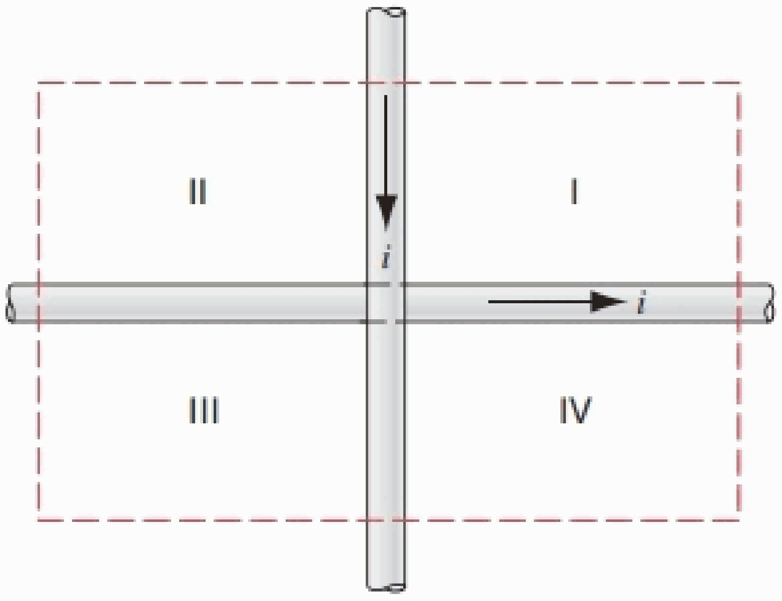


FIGURE 33-33. Question 13.

14. A messy loop of limp wire is placed on a frictionless table and anchored at points a and b as shown in Fig. 33-34. If a current i is now passed through the wire, will it try to form a circular loop or will it try to bunch up further?

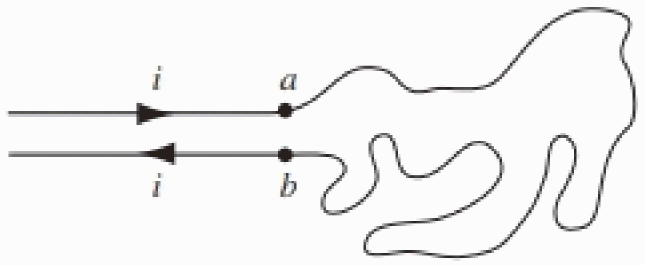


FIGURE 33-34. Question 14.

15. Can the path of integration around which we apply Ampère's law pass through a conductor?

16. Suppose we set up a path of integration around a cable that contains 12 wires with different currents (some in opposite directions) in each wire. How do we calculate i in Ampère's law in such a case?

17. Apply Ampère's law qualitatively to the three paths shown in Fig. 33-35.

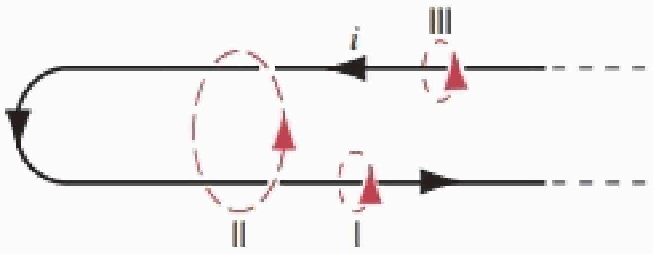


FIGURE 33-35. Question 17.

18. Discuss analogies and differences between Gauss' law and Ampère's law.

19. Does it necessarily follow from symmetry arguments alone that the lines of $\vec{\mathbf{B}}$ around a long, straight wire carrying a current i must be concentric circles?

20. A steady, longitudinal, uniform current is set up in a long copper tube. Is there a magnetic field (*a*) inside and/or (*b*) outside the tube?

21. A very long conductor has a square cross section and contains a coaxial cavity also with a square cross section. Current is distributed uniformly over the material cross section of the conductor. Is the magnetic field in the cavity equal to zero? Justify your answer.

22. A long, straight wire of radius R carries a steady current i. How does the magnetic field generated by this current depend on R? Consider points both outside and inside the wire.

23. A long, straight wire carries a constant current i. What does Ampère's law require for (*a*) a loop that encloses the wire but is not circular, (*b*) a loop that does not enclose the wire, and (*c*) a loop that encloses the wire but does not all lie in one plane?

24. Two long solenoids are nested on the same axis, as in Fig. 33-36. They carry identical currents but in opposite directions. If there is no magnetic field inside the inner solenoid, what can you say about n, the number of turns per unit length, for the two solenoids? Which one, if either, has the larger value?

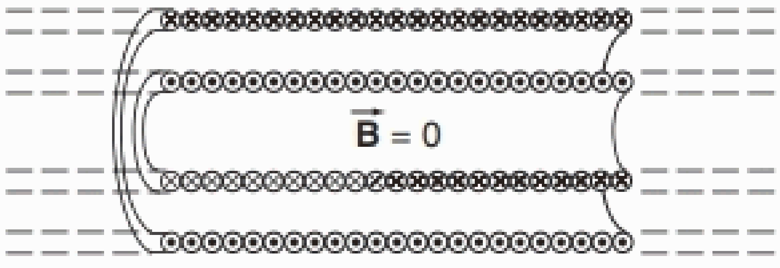


FIGURE 33-36. Question 24.

25. The magnetic field at the center of a circular current loop has the value $B = \mu_0 i/2R$; see Eq. 33-21. However, the electric field at the center of a ring of charge is zero. Why this difference?

26. A steady current is set up in a cubical network of resistive wires, as in Fig. 33-37. Use symmetry arguments to show that the magnetic field at the center of the cube is zero.

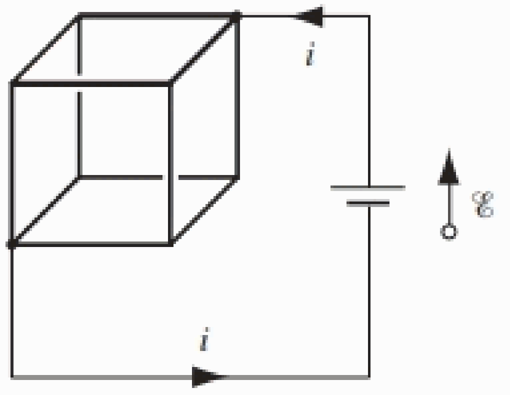


FIGURE 33-37. Question 26.

27. Does Eq. 33-28 ($B = \mu_0 ni$) hold for a solenoid of square cross section?

28. A toroid is described as a solenoid bent into the shape of a doughnut. The magnetic field outside an ideal solenoid is not zero. What can you say about the strength of the magnetic field outside an ideal toroid?

29. Drifting electrons constitute the current in a wire and a magnetic field is associated with this current. What current and magnetic field would be measured by an observer moving along the wire at the electron drift velocity?

EXERCISES

33-1 The Magnetic Field due to a Moving Charge

1. (*a*) What does nonrelativistic physics predict would be the speed of two protons moving side by side a distance d apart so that the magnetic force is exactly balanced by the electric force? (*b*) Comment on the appropriateness of using nonrelativistic expressions for this problem.

33-2 The Magnetic Field of a Current

2. A surveyor is using a magnetic compass 6.3 m below a power line in which there is a steady current of 120 A. Will this interfere seriously with the compass reading? The horizontal component of Earth's magnetic field at the site is 210 μT.

3. A #10 bare copper wire (2.6 mm in diameter) can carry a current of 50 A without overheating. For this current, what is the magnetic field at the surface of the wire?

4. At a location in the Philippines, the Earth's magnetic field has a value of 39.0 μT and is horizontal and due north. The net field is zero 8.13 cm above a long, straight, horizontal wire that carries a steady current. (*a*) Calculate the current and (*b*) find its direction.

5. The 25-kV electron gun in a TV tube fires an electron beam 0.22 mm in diameter at the screen, 5.6×10^{14} electrons arriving each second. Calculate the magnetic field produced by the beam at a point 1.5 mm from the axis of the beam.

6. A straight conductor carrying a current i is split into identical semicircular turns as shown in Fig. 33-38. What is the magnetic field strength at the center C of the circular loop so formed?

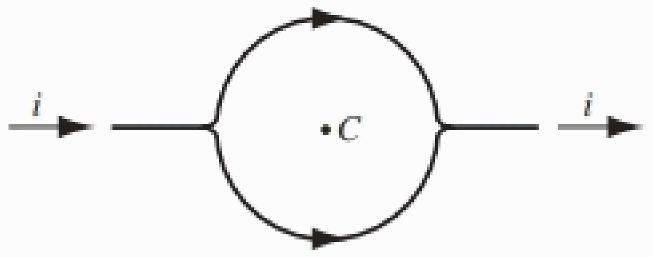


FIGURE 33-38. Exercise 6.

7. A long, straight wire carries a current of 48.8 A. An electron, traveling at 1.08×10^7 m/s, is 5.20 cm from the wire. Calculate the force that acts on the electron if the electron velocity is directed (*a*) toward the wire, (*b*) parallel to the current, and (*c*) at right angles to the directions defined by (*a*) and (*b*).

8. Two long, straight, parallel wires, separated by 0.75 cm, are perpendicular to the plane of the page as shown in Fig. 33-39. Wire W_1 carries a current of 6.6 A into the page. What must be the current (magnitude and direction) in wire W_2 for the resultant magnetic field at point P to be zero?

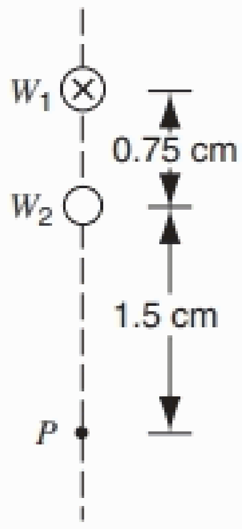


FIGURE 33-39. Exercise 8.

9. Two long, parallel wires are 8.10 cm apart. What equal currents must flow in the wires if the magnetic field halfway between them is to have a magnitude of 296 μT?

10. A long hairpin is formed by bending a piece of wire as shown in Fig. 33-40. If the wire carries a current $i = 11.5$ A, (*a*) what are the magnitude and direction of $\vec{\mathbf{B}}$ at point a? (*b*) At point b, very far from a? Take $R = 5.20$ mm.

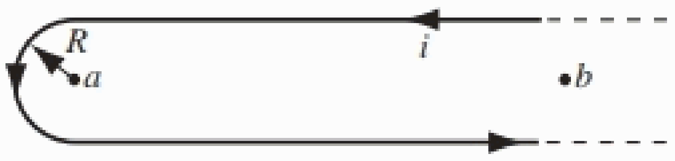


FIGURE 33-40. Exercise 10.

11. A student makes an electromagnet by winding 320 turns of wire around a wooden cylinder of diameter 4.80 cm. The coil is connected to a battery, producing a current of 4.20 A in the wire. At what axial distance $z \gg d$ will the magnetic field of the coil be 5.0 μT (approximately one-tenth the Earth's magnetic field)?

12. A wire carrying current i has the configuration shown in Fig. 33-41. Two semi-infinite straight sections, each tangent to the same circle, are connected by a circular arc, of angle θ, along the circumference of the circle, with all sections lying in the same plane. What must θ be in order for B to be zero at the center of the circle?

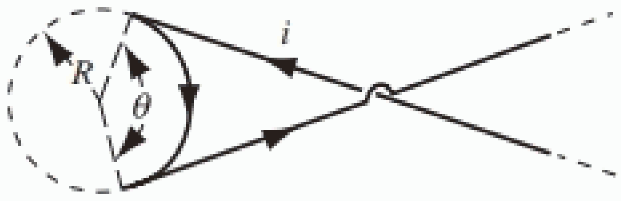


FIGURE 33-41. Exercise 12.

13. Consider the circuit of Fig. 33-42. The curved segments are arcs of circles of radii a and b. The straight segments are along the radii. Find the magnetic field $\vec{\mathbf{B}}$ at P, assuming a current i in the circuit.

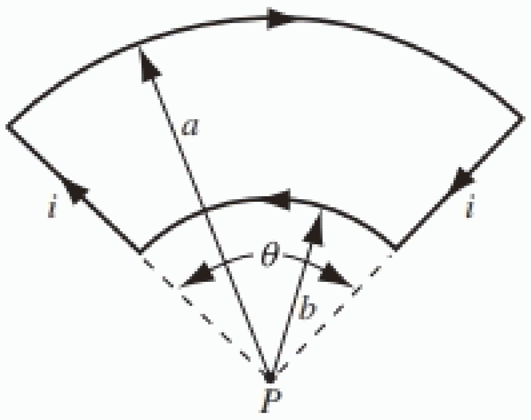


FIGURE 33-42. Exercise 13.

14. Show that B at the center of a rectangular loop of wire of length L and width W, carrying a current i, is given by

$$B = \frac{2\mu_0 i}{\pi} \frac{(L^2 + W^2)^{1/2}}{LW}.$$

Show that this reduces to a result consistent with Sample Problem 33-4 for $L \gg W$.

15. Figure 33-43 shows a cross section of a long, thin ribbon of width w that is carrying a uniformly distributed total current i into the page. Calculate the magnitude and the direction of the magnetic field $\vec{\mathbf{B}}$ at a point P in the plane of the ribbon at a distance d from its edge. (Hint: Imagine the ribbon to be constructed from many long, thin, parallel wires.)

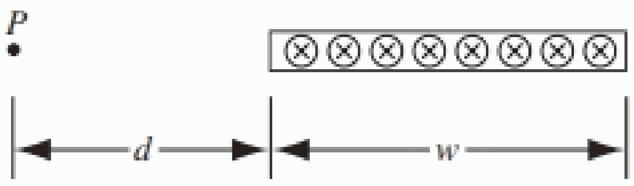


FIGURE 33-43. Exercise 15.

16. Two long, straight, parallel wires 12.2 cm apart each carry a current of 115 A. Figure 33-44 shows a cross section, with the wires running perpendicular to the page and point P lying on the perpendicular bisector of a. Find the magnitude and direction of the magnetic field at P when the current in the left-hand wire is out of the page and the current in the right-hand wire is (a) out of the page and (b) into the page.

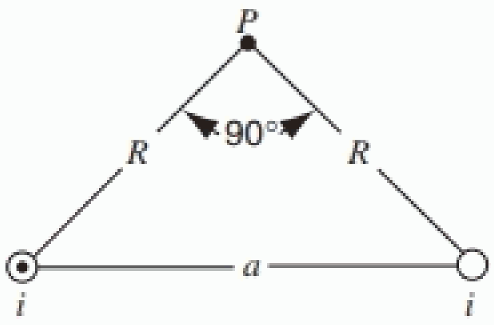


FIGURE 33-44. Exercise 16.

17. In Fig. 33-13a assume that both currents are in the same direction, out of the plane of the figure. Show that the magnetic field in the plane defined by the wires is

$$B = \frac{\mu_0 i x}{\pi(x^2 - b^2)}.$$

Assume that $i = 25$ A and $b = 2.5$ cm in Fig. 33-13a and plot B for the range $-2.5 \text{ cm} < x < +2.5$ cm. Assume that the wire diameters are negligible.

18. Two long wires a distance b apart carry equal antiparallel currents i, as in Fig. 33-45. (a) Show that the magnetic field strength at point P, which is equidistant from the wires, is given by

$$B = \frac{2\mu_0 i b}{\pi(4R^2 + b^2)}.$$

(b) In what direction does $\vec{\mathbf{B}}$ point?

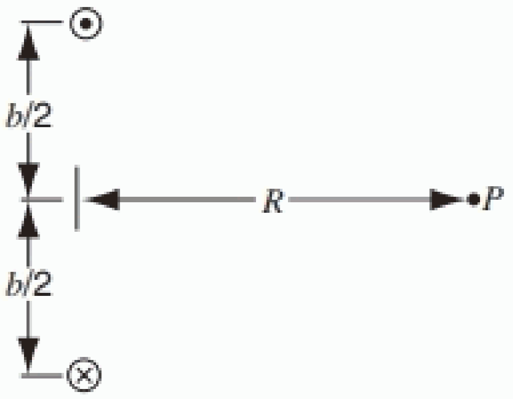


FIGURE 33-45. Exercise 18.

19. A circular loop of radius 12 cm carries a current of 13 A. A second loop of radius 0.82 cm, having 50 turns and a current of 1.3 A, is at the center of the first loop. (a) What magnetic field does the large loop set up at its center? (b) Calculate the torque that acts on the small loop. Assume that the planes of the two loops are at right angles and that the magnetic field due to the large loop is essentially uniform throughout the volume occupied by the small loop.

20. (a) A long wire is bent into the shape shown in Fig. 33-46, without cross contact at P. The radius of the circular section is R. Determine the magnitude and direction of $\vec{\mathbf{B}}$ at the center C of the circular portion when the current i is as indicated. (b) The circular part of the wire is rotated without distortion about the dashed line one quarter turn clockwise as viewed from above, so that the plane of the circular loop is now perpendicular to the plane of the page. Determine $\vec{\mathbf{B}}$ at C in this case.

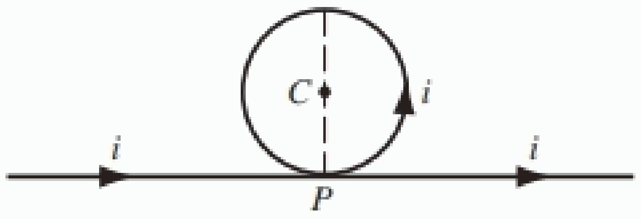


FIGURE 33-46. Exercise 20.

33-3 Two Parallel Currents

21. Four long copper wires are parallel to each other and arranged in a square; see Fig. 33-47. They carry equal currents i out of the page, as shown. Calculate the force per meter on any one wire; give magnitude and direction. Assume that $i = 18.7$ A and $a = 24.5$ cm. (In the case of parallel motion of charged particles in a plasma, this is known as the pinch effect.)

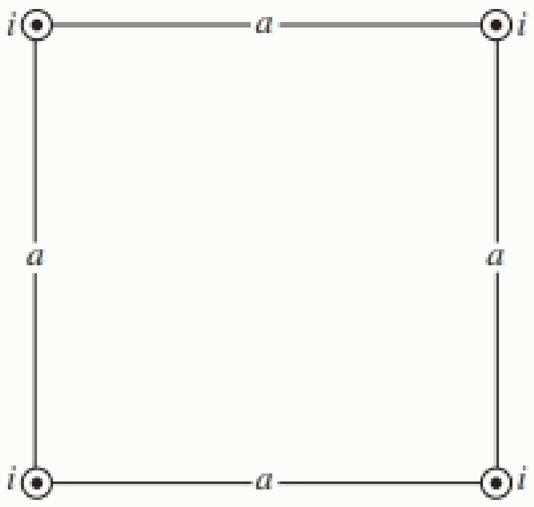


FIGURE 33-47. Exercise 21.

22. Figure 33-48 shows five long, parallel wires in the xy plane. Each wire carries a current $i = 3.22$ A in the positive x direction. The separation between adjacent wires is $a = 8.30$ cm. Find the magnetic force per meter, magnitude and direction, exerted on each of these five wires.

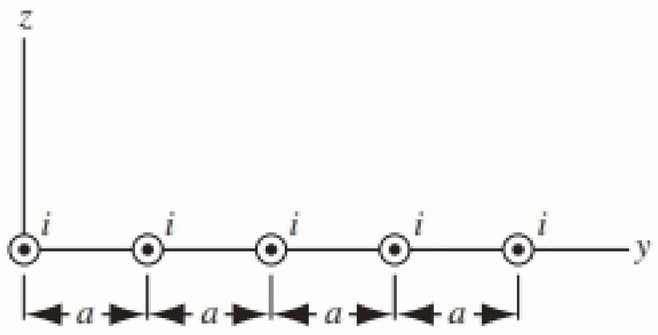


FIGURE 33-48. Exercise 22.

23. Figure 33-49 shows an idealized schematic of an "electromagnetic rail gun," which is designed to fire projectiles at speeds up to 10 km/s. The projectile P sits between and in contact with two parallel rails along which it can slide. A generator G provides a current that flows up one rail, across the projectile, and back down the other rail. (a) Let w be the distance between the rails, r the radius of the rails (presumed circular), and i the current. Show that the force on the projectile is to the right and given approximately by

$$F = \frac{1}{2}\left(\frac{i^2\mu_0}{\pi}\right)\ln\left(\frac{w+r}{r}\right).$$

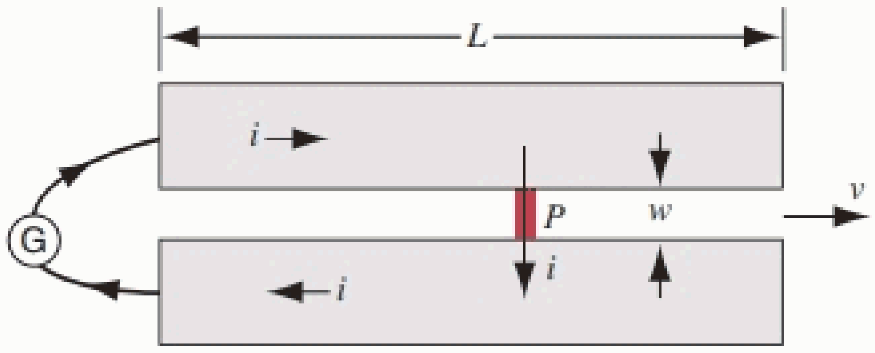


FIGURE 33-49. Exercise 23.

(b) If the projectile starts from the left end of the rail at rest, find the speed v at which it is expelled at the right. Assume that $i = 450$ kA, $w = 12$ mm, $r = 6.7$ cm, $L = 4.0$ m, and that the mass of the projectile is $m = 10$ g.

24. Figure 33-50 shows a long wire carrying a current i_1. The rectangular loop carries a current i_2. Calculate the resultant force acting on the loop. Assume that $a = 1.10$ cm, $b = 9.20$ cm, $L = 32.3$ cm, $i_1 = 28.6$ A, and $i_2 = 21.8$ A.

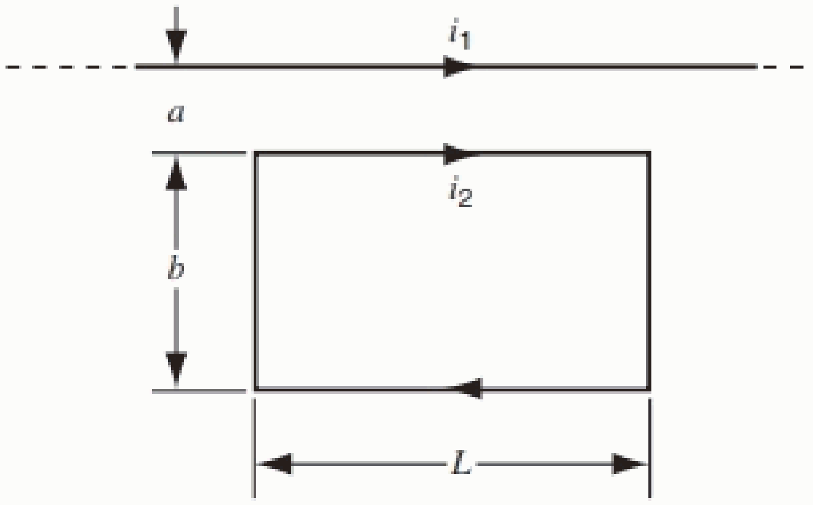


FIGURE 33-50. Exercise 24.

25. In Sample Problem 33-6, suppose that the upper wire is displaced downward a small distance and then released. Show that the resulting motion of the wire is simple harmonic with the same frequency of oscillation as a simple pendulum of length d.

33-4 The Magnetic Field of a Solenoid

26. A solenoid 95.6 cm long has a radius of 1.90 cm, a winding of 1230 turns, and carries a current of 3.58 A. Calculate the strength of the magnetic field inside the solenoid.

27. A solenoid 1.33 m long and 2.60 cm in diameter carries a current of 17.8 A. The magnetic field inside the solenoid is 22.4 mT. Find the length of the wire forming the solenoid.

28. A long solenoid with 115 turns/cm and a radius of 7.20 cm carries a current of 1.94 mA. A current of 6.30 A flows in a straight conductor along the axis of the solenoid. (a) At what radial distance from the axis will the direction of the resulting magnetic field be at 40.0° from the axial direction? (b) What is the magnitude of the magnetic field?

29. Verify the integration in Eq. 33-27, and show that as $L \rightarrow \infty$, Eq. 33-27 approaches Eq. 33-28.

33-5 Ampère's Law

30. Eight wires cut the page perpendicularly at the points shown in Fig. 33-51. A wire labeled with the integer k ($k = 1, 2, \ldots, 8$) bears the current ki_0. For those with odd k, the current is out of the page; for those with even k it is into the page. Evaluate $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}}$ along the closed loop in the direction shown.

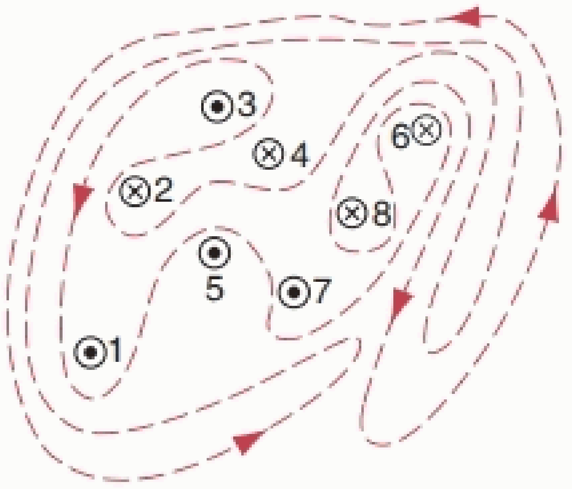


FIGURE 33-51. Exercise 30.

31. Each of the indicated eight conductors in Fig. 33-52 carries 2.0 A of current into or out of the page. Two paths are indicated for the line integral $\vec{\mathbf{B}} \cdot d\vec{s}$. What is the value of the integral for (*a*) the dotted path and (*b*) the dashed path?

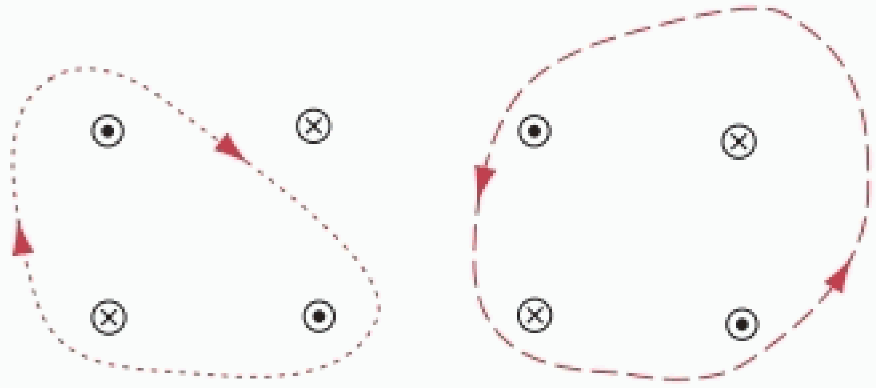


FIGURE 33-52. Exercise 31.

32. Consider a long, cylindrical wire of radius R carrying a current i distributed uniformly over the cross section. At what two distances from the axis of the wire is the magnetic field strength, due to the current, equal to one-half the value at the surface?

33. Figure 33-53 shows a cross section of a long conductor of a type called a coaxial cable of radii a, b, and c. Equal but antiparallel, uniformly distributed currents i exist in the two conductors. Derive expressions for $B(r)$ in the ranges (*a*) $r < c$, (*b*) $c < r < b$, (*c*) $b < r < a$, and (*d*) $r > a$. (*e*) Test these expressions for all the special cases that occur to you. (*f*) Assume that $a = 2.0$ cm, $b = 1.8$ cm, $c = 0.40$ cm, and $i = 120$ A and plot $B(r)$ over the range $0 < r < 3$ cm.

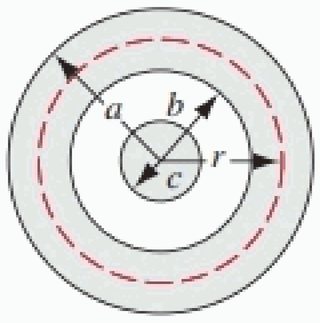


FIGURE 33-53. Exercise 33.

34. Figure 33-54 shows a cross section of a hollow, cylindrical conductor of radii a and b, carrying a uniformly distributed current i. (*a*) Using the circular Ampèrian loop shown, verify that $B(r)$ for the range $b < r < a$ is given by

$$B(r) = \frac{\mu_0 i}{2\pi(a^2 - b^2)} \frac{r^2 - b^2}{r}.$$

(*b*) Test this formula for the special cases of $r = a$, $r = b$, and $b = 0$. (*c*) Assume that $a = 2.0$ cm, $b = 1.8$ cm, and $i = 100$ A and plot $B(r)$ for the range $0 < r < 6$ cm.

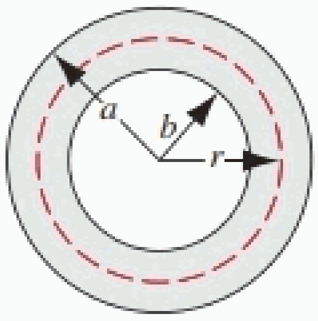


FIGURE 33-54. Exercise 34.

35. A long, circular pipe, with an outside radius of R, carries a (uniformly distributed) current of i_0 (into the paper as shown in Fig. 33-55). A wire runs parallel to the pipe at a distance $3R$ from center to center. Calculate the magnitude and direction of the current in the wire that would cause the resultant magnetic field at the point P to have the same magnitude, but the opposite direction, as the resultant field at the center of the pipe.

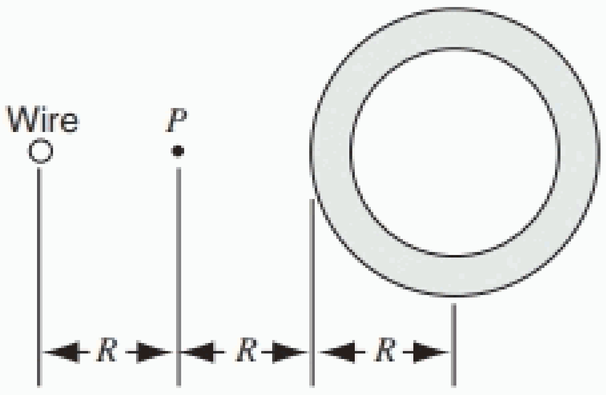


FIGURE 33-55. Exercise 35.

36. A toroid having a square cross section, 5.20 cm on edge, and an inner radius of 16.2 cm has 535 turns and carries a current of 813 mA. Calculate the magnetic field inside the toroid at (*a*) the inner radius and (*b*) the outer radius of the toroid.

37. An interesting (and frustrating) effect occurs when one attempts to confine a collection of electrons and positive ions (a plasma) in the magnetic field of a toroid. Particles whose motion is perpendicular to the $\vec{\mathbf{B}}$ field will not execute circular paths because the field strength varies with radial distance from the axis of the toroid. This effect, which is shown (exaggerated) in Fig. 33-56, causes particles of opposite sign to drift in opposite directions parallel to the axis of the toroid. (*a*) What is the sign of the charge on the particle whose path is sketched in the figure? (*b*) If the particle path has a radius of curvature of 11 cm when its radial distance from the axis of the toroid is 125 cm, what will be the radius of curvature when the particle is 110 cm from the axis?

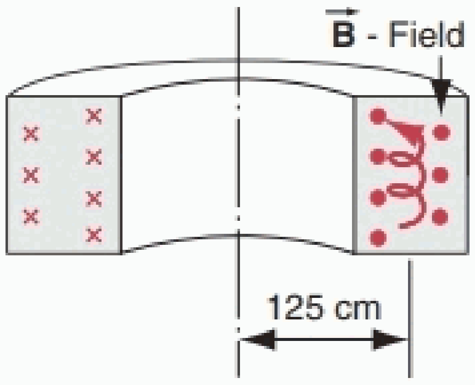


FIGURE 33-56. Exercise 37.

33-6 Electromagnetism and Frames of Reference

PROBLEMS

1. Figure 33-57 shows an arrangement known as a *Helmholtz coil*. It consists of two circular coaxial coils each of N turns and radius R, separated by a distance R. They carry equal currents i in the same direction. Find the magnetic field at P, midway between the coils.

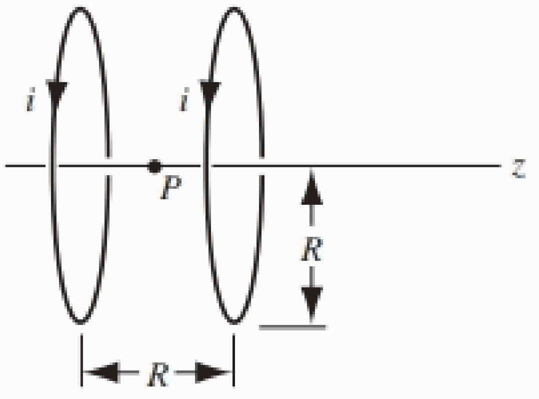


FIGURE 33-57. Problems 1, 3, and computer problem 1.

2. A straight section of wire of length L carries a current i. (*a*) Show that the magnetic field associated with this segment at P, a perpendicular distance D from one end of the wire (see Fig. 33-58), is given by

$$B = \frac{\mu_0 i}{4\pi D}\frac{L}{(L^2 + D^2)^{1/2}}.$$

(*b*) Show that the magnetic field is zero at point Q, along the line of the wire.

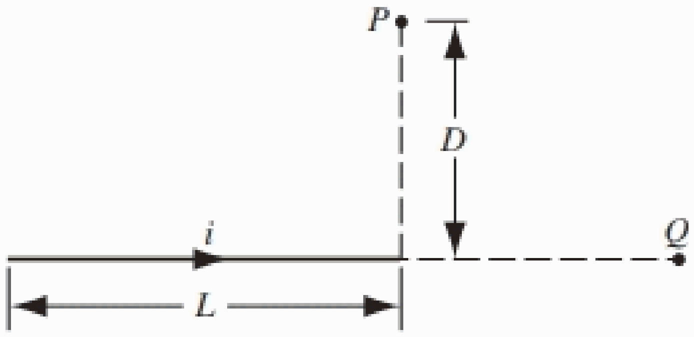


FIGURE 33-58. Problem 2.

3. In Problem 1 (Fig. 33-57) let the separation of the coils be a variable s (not necessarily equal to the coil radius R). (*a*) Show that the first derivative of the magnetic field (dB/dz) vanishes at the midpoint P regardless of the value of s. Why would you expect this to be true from symmetry? (*b*) Show that the second derivative of the magnetic field (d^2B/dz^2) also vanishes at P if $s = R$. This accounts for the uniformity of B near P for this particular coil separation.

4. A square loop of wire of edge a carries a current i. (*a*) Show that B for a point on the axis of the loop and a distance z from its center is given by

$$B(z) = \frac{4\mu_0 i a^2}{\pi(4z^2 + a^2)(4z^2 + 2a^2)^{1/2}}.$$

(*b*) To what does this reduce at the center of the loop?

5. (*a*) A wire in the form of a regular polygon of n sides is just enclosed by a circle of radius a. If the current in this wire is i, show that the magnetic field $\vec{\mathbf{B}}$ at the center of the circle is given in magnitude by

$$B = \frac{\mu_0 n i}{2\pi a}\tan(\pi/n).$$

(*b*) Show that as $n \to \infty$, this result approaches that of a circular loop.

6. You are given a length L of wire in which a current i may be established. The wire may be formed into a circle or a square. Show that the square yields the greater value for B at the central point.

7. (*a*) Calculate $\vec{\mathbf{B}}$ at point P in Fig. 33-59. (*b*) Is the field strength at P greater or less than at the center of the square?

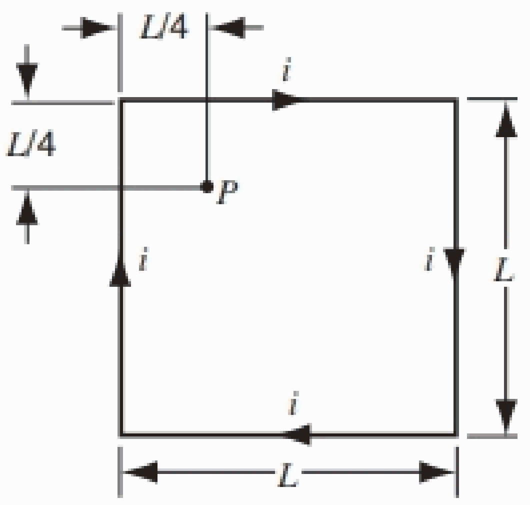


FIGURE 33-59. Problem 7.

8. A thin plastic disk of radius R has a charge q uniformly distributed over its surface. If the disk rotates at an angular frequency ω about its axis, show that the magnetic field at the center of the disk is

$$B = \frac{\mu_0 \omega q}{2\pi R}.$$

(Hint: The rotating disk is equivalent to an array of current loops.)

9. A long solenoid has 100 turns per centimeter. An electron moves within the solenoid in a circle of radius 2.30 cm perpendicular to the solenoid axis. The speed of the electron is $0.0460c$ (c = speed of light). Find the current in the solenoid.

10. In a certain region there is a uniform current density of 15 A/m^2 in the positive z direction. What is the value of $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}}$ when the line integral is taken along the three straight-line segments from $(4d, 0, 0)$ to $(4d, 3d, 0)$ to $(0, 0, 0)$ to $(4d, 0, 0)$, where $d = 23$ cm?

11. Show that a uniform magnetic field $\vec{\mathbf{B}}$ cannot drop abruptly to zero as one moves at right angles to it, as suggested by the horizontal arrow through point a in Fig. 33-60. (Hint: Apply Ampère's law to the rectangular path shown by the dashed lines.) In actual magnets, "fringing" of the lines of $\vec{\mathbf{B}}$ always occurs, which means that $\vec{\mathbf{B}}$ approaches zero in a gradual manner. Modify the $\vec{\mathbf{B}}$ lines in the figure to indicate a more realistic situation.

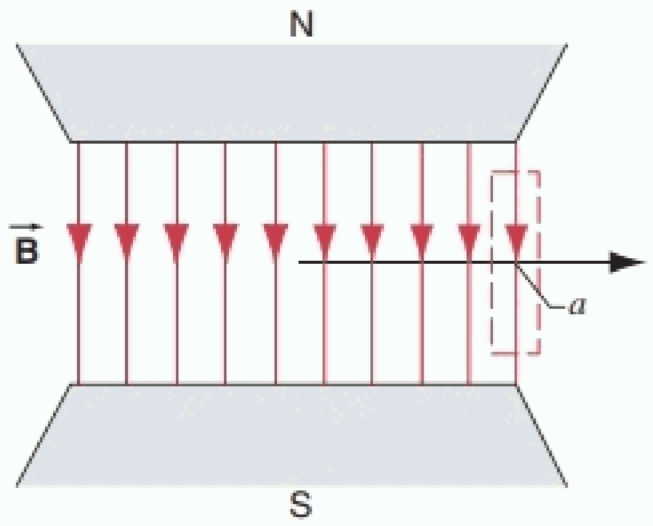


FIGURE 33-60. Problem 11.

12. A conductor consists of an infinite number of adjacent wires, each infinitely long and carrying a current i. Show that the lines of $\vec{\mathbf{B}}$ are as represented in Fig. 33-61 and that B for all points above and below the infinite current sheet is given by

$$B = \tfrac{1}{2}\mu_0 n i,$$

where n is the number of wires per unit length. Derive both by direct application of Ampère's law and by considering the problem as a limiting case of Sample Problem 33-5.

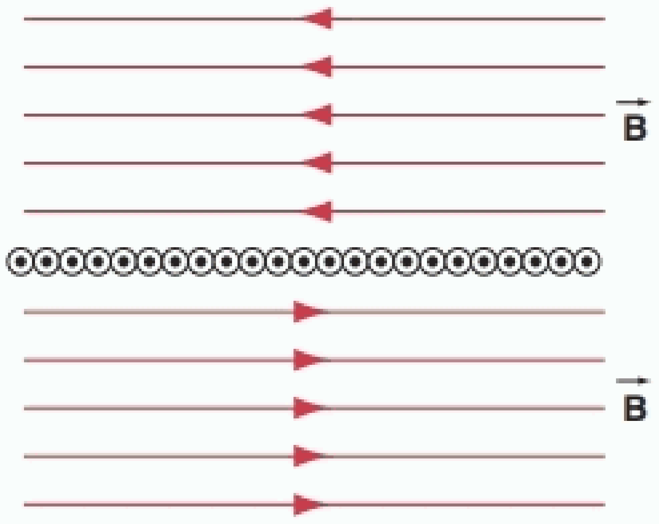


FIGURE 33-61. Problem 12.

13. The current density inside a long, solid, cylindrical wire of radius a is in the direction of the axis and varies linearly with radial distance r from the axis according to $j = j_0 r/a$. Find the magnetic field inside the wire. Express your answer in terms of the total current i carried by the wire.

14. Figure 33-62 shows a cross section of a long, cylindrical conductor of radius R containing a long, cylindrical hole of radius a. The axes of the two cylinders are parallel and are a distance b apart. A current i is uniformly distributed over the shaded area in the figure. (*a*) Use superposition ideas to show that the magnetic field at the center of the hole is

$$B = \frac{\mu_0 i b}{2\pi(R^2 - a^2)}.$$

(*b*) Discuss the two special cases $a = 0$ and $b = 0$. (*c*) Can you use Ampère's law to show that the magnetic field in the hole is uniform? (Hint: Regard the cylindrical hole as filled with two equal currents moving in opposite directions, thus canceling each other. Assume that each of these currents has the same current density as that in the actual conductor. Thus we superimpose the fields due to two complete cylinders of current, of radii R and a, each cylinder having the same current density.)

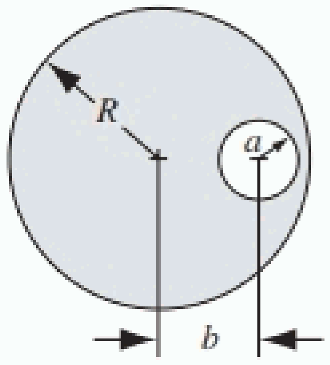


FIGURE 33-62. Problem 14.

COMPUTER PROBLEMS

1. Two 300-turn coils each carry a current i. They are arranged a distance apart equal to their radius, as in Fig. 33-57. (This is the Helmholtz coil geometry; see Problem 1.) For $R = 5.0$ cm and $i = 50$ A, plot B as a function of distance z along the common axis over the range $z = -5$ cm to $z = +5$ cm, taking $z = 0$ at the midpoint P. Such coils provide an especially uniform field B near point P.

2. Design a double Helmholtz coil so that d^4B/dz^4 also vanishes at the center. This is a problem best solved on a computer program such as Mathematica or MAPLE.

CHAPTER 34

FARADAY'S LAW OF INDUCTION

We can often anticipate the outcome of an experiment by considering how it is related by symmetry to other experiments. For example, a current loop in a magnetic field experiences a torque (due to the field) that rotates the loop. Consider a similar situation: a loop of wire in which there is no current is placed in a magnetic field, and a torque applied by an external agent rotates the loop. We find that a current appears in the loop! For a loop of wire in a magnetic field, a current produces a torque, and a torque produces a current. This is an example of the symmetry of nature.

We can understand and analyze the appearance of current in the loop based on Faraday's law of induction, *which is the subject of this chapter. Faraday's law, which is one of the four Maxwell equations, can be directly demonstrated from a number of simple experiments that can easily be done in the laboratory.*

34-1 FARADAY'S EXPERIMENTS

Faraday's law of induction was discovered through experiments carried out by Michael Faraday in England in 1831 and by Joseph Henry in the United States at about the same time.* Even though Faraday published his results first, which gives him priority of discovery, the SI unit of inductance (see Chapter 36) is called the *henry* (abbreviation H). On the other hand, the SI unit of capacitance is, as we have seen, called the *farad* (abbreviation F). In Chapter 36, where we discuss oscillations in capacitative–inductive circuits, we see how appropriate it is to link the names of these two talented contemporaries in a single context.

Figure 34-1 shows a coil of wire as a part of a circuit containing an ammeter. Normally, we would expect the ammeter to show no current in the circuit because there seems to be no electromotive force. However, if we push a bar magnet toward the coil, with its north pole facing the coil, a remarkable thing happens. *While the magnet is moving,* the ammeter deflects, showing that a current has been set up in the coil. If we hold the magnet stationary with respect to the coil, the ammeter does not deflect. If we move the

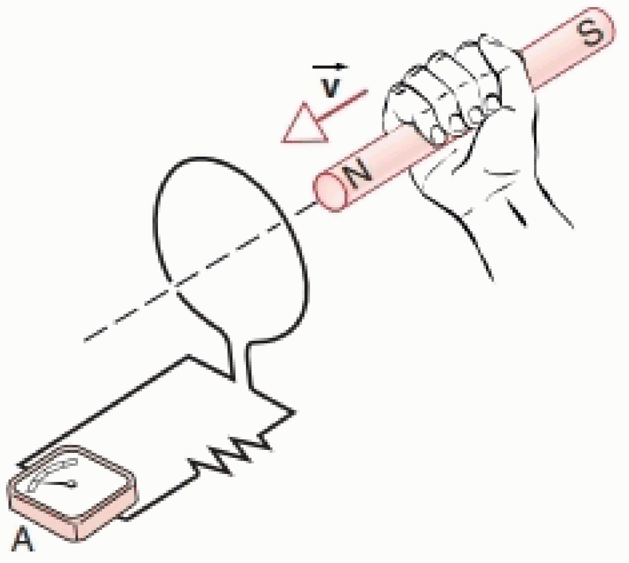


FIGURE 34-1. The ammeter A deflects, indicating a current in the circuit, when the magnet is moving with respect to the coil.

---

*In addition to their independent simultaneous discovery of the law of induction, Faraday and Henry had several other similarities in their lives. Both were apprentices at an early age. Faraday, at age 14, was apprenticed to a London bookbinder. Henry, at age 13, was apprenticed to a watchmaker in Albany, New York. In later years Faraday was appointed director of the Royal Institution in London, whose founding was due in large part to an American, Benjamin Thompson (Count Rumford). Henry, on the other hand, became secretary of the Smithsonian Institution in Washington, DC, which was founded by an endowment from an Englishman, James Smithson.

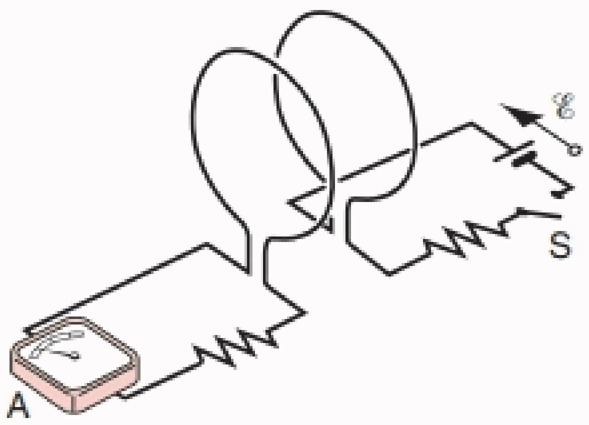


FIGURE 34-2. The ammeter A deflects momentarily when switch S is closed or opened. No physical motion of the coils is involved.

magnet *away* from the coil, the meter again deflects, but in the opposite direction, which means that the current in the coil is in the opposite direction. If we use the south pole end of a magnet instead of the north pole end, the experiment works as described but the deflections are reversed. The faster the magnet is moved, the greater is the reading of the meter. Further experimentation shows that *what matters is the relative motion of the magnet and the coil.* It makes no difference whether we move the magnet toward the coil or the coil toward the magnet.

The current that appears in this experiment is called an *induced current* and is said to be set up by an *induced electromotive* force. Note that there are no batteries anywhere in the circuit. From experiments like this, Faraday was able to deduce the law that gives the magnitude and direction of the induced emfs. Such emfs are very important in practice. The chances are good that the lights in the room in which you are reading this book are operated from an induced emf produced in a commercial electric generator.

In another experiment, the apparatus of Fig. 34-2 is used. The coils are placed close together but at rest with respect to each other. When we close the switch S, thus setting up a steady current in the right-hand coil, the meter in the left-hand coil deflects momentarily. When we open the switch, thus interrupting this current, the meter again deflects momentarily, but in the opposite direction. None of the apparatus is physically moving in this experiment.

Experiment shows that there is an induced emf in the left coil of Fig. 34-2 whenever the current in the right coil is *changing*. It is the *rate at which the current is changing and not the size of the current* that is significant.

The common feature of these two experiments is *motion* or *change*. It is the *moving* magnet or the *changing* current that is responsible for the induced emfs. In the next section, we give the mathematical basis for these effects.

34-2 FARADAY'S LAW OF INDUCTION

Imagine that there are lines of magnetic field coming from the bar magnet of Fig. 34-1 and from the right-hand current loop in Fig. 34-2. Some of those field lines pass through the left-hand coil in both figures. As the magnet is moved in the situation of Fig. 34-1, or as the switch is opened or closed in Fig. 34-2, the number of lines of the magnetic field passing through the left-hand coil changes. As Faraday's experiments showed and as Faraday's technique of field lines helps us visualize, *it is the change in the number of field lines passing through a circuit loop that induces the emf in the loop.* Specifically, it is the *rate of change* in the number of field lines passing through the loop that determines the induced emf.

To make this statement quantitative, we introduce the *magnetic flux* Φ_B. Like the electric flux (see Section 27-3), the magnetic flux can be considered to be a measure of the number of field lines passing through a surface. In analogy with the electric flux (see Eq. 27-7), the magnetic flux through *any* surface is defined as

$$\Phi_B = \int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}}. \tag{34-1}$$

Here $d\vec{\mathbf{A}}$ is an element of area of the surface (shown in Fig. 34-3), and the integration is carried out over the entire surface through which we wish to calculate the flux (for example, the surface enclosed by the loop in Fig. 34-1). If the magnetic field has a constant magnitude and direction over a planar area A, the flux can be written

$$\Phi_B = BA\cos\theta, \tag{34-2}$$

where θ is the angle between the normal to the surface and the direction of the field.

The SI unit of magnetic flux is the tesla · meter², which is given the name of *weber* (abbreviation Wb); that is,

$$1 \text{ weber} = 1 \text{ tesla} \cdot \text{meter}^2.$$

Inverting this relationship, we see that the tesla is equivalent to the weber/meter², which was the unit used for magnetic fields before the tesla was adopted as the SI unit.

In terms of the magnetic flux, the emf induced in a circuit is given by *Faraday's law of induction:*

> *The magnitude of the induced emf in a circuit is equal to the rate at which the magnetic flux through the circuit is changing with time.*

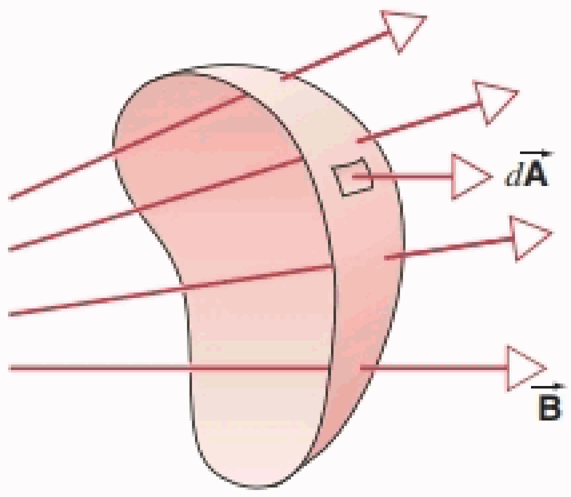


FIGURE 34-3. The magnetic field $\vec{\mathbf{B}}$ through an area A gives a magnetic flux through the surface. The element of area $d\vec{\mathbf{A}}$ is represented by a vector.

In mathematical terms, Faraday's law is

$$|\mathcal{E}| = \left|\frac{d\Phi_B}{dt}\right|, \tag{34-3}$$

where $\mathcal{E}$ is the induced emf. If the rate of change of flux is in units of webers per second, the emf has units of volts. We consider the direction (or sign) of the emf in the next section.

If the coil consists of N turns, then an induced emf appears in every turn, and the total induced emf in the circuit is the sum of the individual values, just as in the case of batteries connected in series. If the coil is so tightly wound that each turn may be considered to occupy the same region of space and therefore to experience the same change of flux, then the total induced emf is

$$|\mathcal{E}| = N\left|\frac{d\Phi_B}{dt}\right|. \tag{34-4}$$

There are many ways of changing the flux through a loop: moving a magnet relative to the loop (as in Fig. 34-1), changing the current in a nearby circuit (as in Fig. 34-2 and also as in a transformer), moving the loop in a nonuniform field, rotating the loop in a fixed magnetic field such that the angle θ in Eq. 34-2 changes (as in a generator, which we discuss in Section 34-5), or changing the size or shape of the loop. In each of these methods, an emf is induced in the loop.

Finally, we note that, even though Eq. 34-3 is known as Faraday's law, it was not written in that form by Faraday, who was untrained in mathematics. In fact, Faraday's three-volume published work on electromagnetism, a landmark achievement in the development of physics and chemistry, contains not a single equation!

SAMPLE PROBLEM 34-1. The long solenoid S of Fig. 34-4 has 220 turns/cm; its diameter d is 3.2 cm. At its center we place a 130-turn close-packed coil C of diameter $d_C = 2.1$ cm. The current in the solenoid is increased from zero to 1.5 A at a steady rate over a period of 0.16 s. What is the absolute value (that is, the magnitude without regard for sign) of the induced emf that appears in the central coil while the current in the solenoid is being changed?

Solution In this case the flux through the coil C is changing because the field B in the interior of the solenoid varies as the current changes. The field is uniform and perpendicular to the area of each turn of coil C, so $\Phi_B = \int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = BA$, where A is the area of each turn of the inner coil. Faraday's law gives

$$|\mathcal{E}| = N\left|\frac{\Delta\Phi_B}{\Delta t}\right| = N\left|\frac{A\,\Delta B}{\Delta t}\right| = NA\left|\frac{\mu_0 n\,\Delta i}{\Delta t}\right| = NA\mu_0 n\left|\frac{\Delta i}{\Delta t}\right|,$$

where we have used $B = \mu_0 ni$ for the field of an ideal solenoid (Eq. 33-28). The area A of the central coil is $\frac{1}{4}\pi d_C^2 = 3.46 \times 10^{-4}\ \mathrm{m^2}$. The current changes by 1.5 A in 0.16 s. The emf is then

$$\begin{aligned}|\mathcal{E}| &= (130)(3.46 \times 10^{-4}\ \mathrm{m^2})(4\pi \times 10^{-7}\ \mathrm{T\cdot m/A})\\ &\quad \times (2.2 \times 10^4\ \mathrm{turns/m})\ \frac{1.5\ \mathrm{A}}{0.16\ \mathrm{s}}\\ &= 1.2 \times 10^{-2}\ \mathrm{V} = 12\ \mathrm{mV}.\end{aligned}$$

We shall explain in the next section how to find the *direction* of the induced emf. We can preview that discussion with the following argument. Suppose an increase in the flux from the outer coil caused a current in the inner coil that produced a magnetic field in the same direction as the original field. This would in turn increase the flux through the area enclosed by the outer coil, which should similarly cause its current to increase, thereby increasing again the current in the inner coil, and so on. Is this a reasonable outcome?

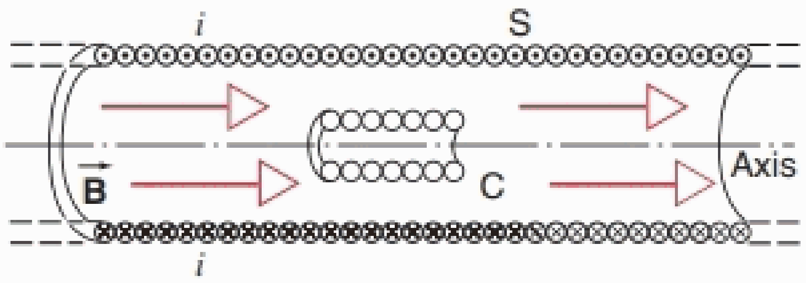


FIGURE 34-4. Sample Problem 34-1. A coil C is located inside a solenoid S. The solenoid carries a current that emerges from the page at the top and enters at the bottom, as indicated by the dots and crosses. When the current in the solenoid is changing, an induced emf appears in the coil.

34-3 LENZ' LAW

So far we have not specified the direction of the induced emf. Later in this chapter we discuss the relationship between induced emf and induced electric field, which allows us to write Faraday's law without the absolute magnitude bars. For now, we determine the direction of the induced emf based on the (induced) current that it would produce, using a rule proposed in 1834 by Heinrich Lenz (1804–1865) and known as *Lenz' law:*

> *The flux of the magnetic field due to the induced current opposes the change in flux that causes the induced current.*

Lenz' law refers to induced *currents,* which means that it applies only to closed conducting circuits. If the circuit is open, we can usually think in terms of what would happen if it *were* closed and in this way find the direction of the induced emf.

If the "change in flux" is an *increase,* then Lenz' law requires that the direction of the induced current oppose the increase; that is, the flux from the magnetic field of the induced current would pass through the loop in a direction opposite to that of the original flux that is increasing. If the "change in flux" is a *decrease,* the flux from the magnetic field of the induced current opposes the decrease; that is, it tends to add to the original flux to keep it from decreasing.

As an example of the application of Lenz' law, let us consider the first of Faraday's experiments shown in Fig. 34-1. As the north pole of the magnet is moved toward the

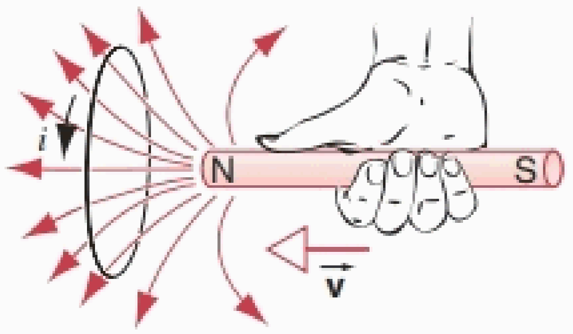


FIGURE 34-5. When the magnet is pushed toward the loop, the magnetic flux through the loop is increased. The induced current through the loop sets up a magnetic field that opposes the increase in flux. The field of the current loop is not shown.

loop, the flux through the loop increases. This increase in flux is the "change" to which Lenz' law refers. The current in the loop must oppose this change; that is, the current must set up a magnetic field that points in the opposite direction inside the loop. Figure 34-5 shows the direction of the induced current. Using the right-hand rule, you can see that inside the loop the magnetic field set up by the induced current is in a direction opposite to that of the magnet.

If, on the other hand, we moved the magnet *away from* the loop, the flux would decrease, and the induced current must oppose this decrease. Therefore, the magnetic field due to the induced current must add to the field of the magnet inside the loop, so the current would be in the opposite direction to that shown in Fig. 34-5.

Another way to interpret Lenz' law is based on the principle of conservation of energy. When the magnet is moved toward the loop, the field due to the induced current in the loop exerts a force that opposes the motion of the magnet, as indicated in Fig. 34-6. The loop sets up a magnetic field similar to that of a magnet with its north pole toward the north pole of the approaching magnet, and the two north poles repel one another. That is, we must exert a greater force to continue pushing the magnet toward the loop. Suppose instead that the current in the loop were in the opposite direction, so that it would set up an induced field in the opposite direction. Rather than being repelled by the field of the induced current, the magnet would be attracted by this field and would be accelerated toward the loop. As the magnet accelerates, the current in the loop would increase, causing an increasing force on the magnet and an increasing acceleration. Both the kinetic energy of the magnet and the rate of Joule (i^2R) energy dissipation in the loop would increase. As a result of a small initial push of the magnet toward the loop, we would get a large increase in energy—clearly a violation of conservation of energy. This cannot occur, and so we conclude that the field due to the induced current must oppose the motion of the magnet toward the loop. Similarly, if we were to move the magnet away from the loop, the induced field would oppose this motion and would pull the magnet toward the loop.

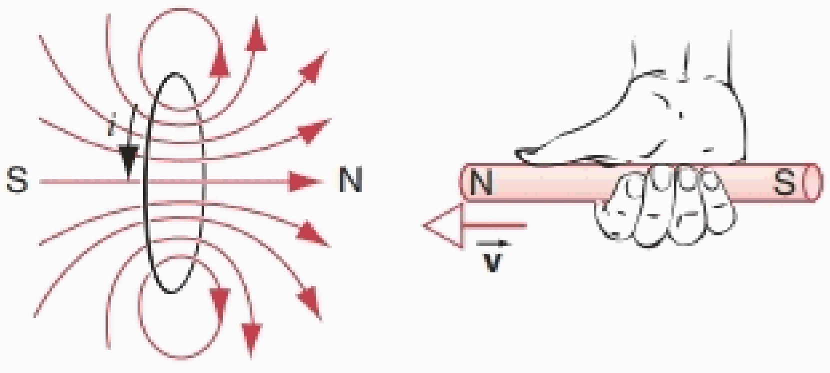


FIGURE 34-6. When the magnet is pushed toward the loop, the induced current *i* has the direction shown, setting up a magnetic field that opposes the motion of the magnet. The field of the magnet is not shown here, but is the same as Fig. 34-5.

In either case, it is not significant that the induced field opposes the *field* of the magnet, but rather that it opposes the *change,* which is an increase or decrease of flux through the loop. If in Fig. 34-5 the field of the magnet points to the left and is increasing (as the magnet moves toward the loop), the induced field must point to the right inside the loop. If the field of the magnet points to the left and is decreasing (as the magnet moves away from the loop), the induced field must point to the left inside the loop. If the magnet were turned around so that the south pole moved toward the loop, the field would point to the right and increase, so that the induced field would point to the left inside the loop.

We can now obtain the direction of the current in the small coil C of Sample Problem 34-1. The field of the solenoid S points to the right in Fig. 34-4 and is increasing. The induced current in C must oppose this increase in flux through C and so must set up a field that opposes the field of S. The current in C is therefore in a direction opposite to that in S. If the current in S were *decreasing* instead of increasing, a similar argument shows that the induced current in C would have the same direction as the current in S.

Signs in Faraday's Law

So far we have written Faraday's law in terms of magnitudes only. The direction and magnitude of the induced current in a conducting loop can be determined using Lenz' law and the magnitude-only version of Faraday's law.

We would like now to remove the absolute magnitude bars from Faraday's law (Eq. 34-3). Before we do so, we must clarify ambiguities in sign that exist on *both* sides of the equation. If we want to write $\mathcal{E}$ instead of $|\mathcal{E}|$, we must specify what we mean by the sign or the direction of an induced emf. Consider the simple one-loop circuit of Fig. 34-7*a*, which may include several batteries and resistors. To analyze this circuit we could use the loop rule. It should be clear that if we travel around clockwise, add up all the emfs, and find a net positive value, then the current will also be clockwise. (Only in that way can the total potential differences across the resistors be negative, in order to satisfy the loop theorem.) For real circuits, the direction of positive current is the same as the direction of travel around the loop in which we would obtain a net positive emf. We can make the same connection for induced currents and emfs: *the direction of the induced current is the direction around the loop that we travel to obtain a positive emf.*

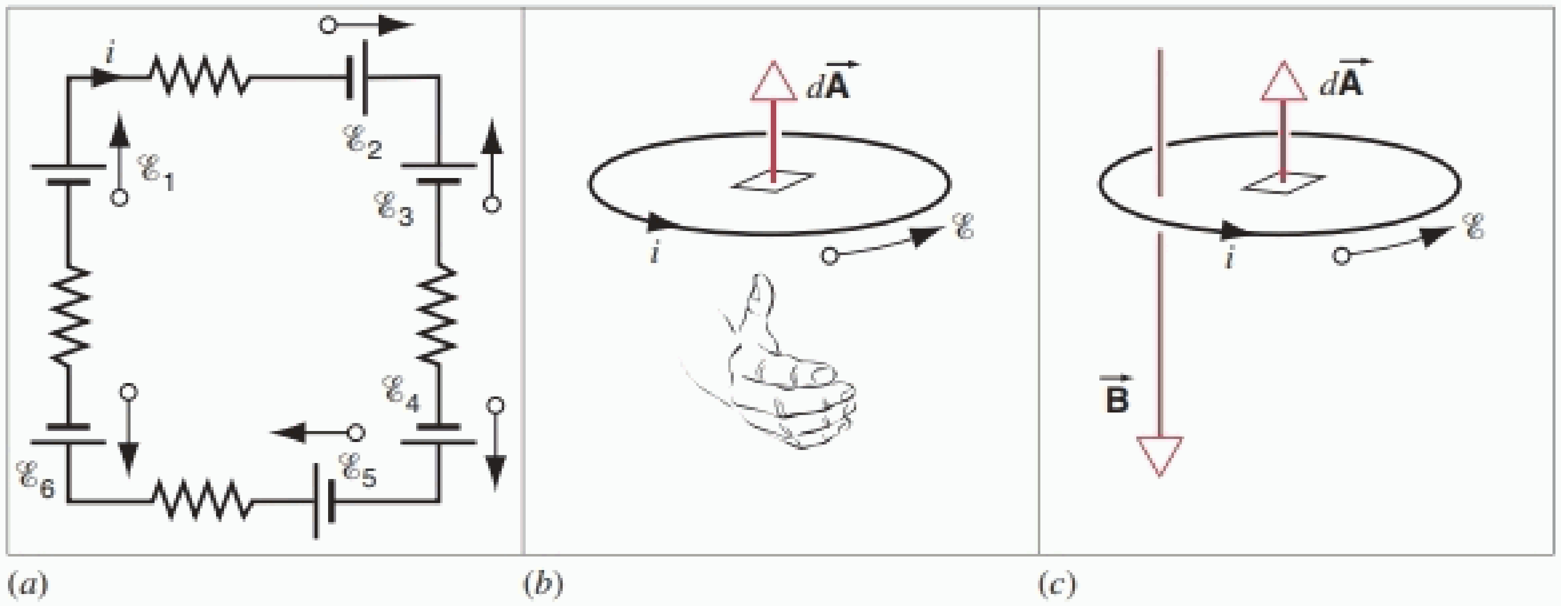


FIGURE 34-7. (*a*) Adding the emfs in a circuit loop to find the direction of the current. (*b*) The right-hand rule for Faraday's law: with the fingers in the direction of i, the thumb is in the direction of $d\vec{\mathbf{A}}$. (*c*) When $\vec{\mathbf{B}}$ is downward and increasing in magnitude, the induced current is as shown.

The sign of the right side of Eq. 34-3 also presents us with a difficulty. When we discussed Gauss' law, which involved the flux through a *closed* surface, we defined the direction of $d\vec{\mathbf{A}}$ to be that of the outward normal through the closed surface. However, the surfaces bounded by the current loops to which we apply Faraday's law are open surfaces, and so it appears that we could choose $d\vec{\mathbf{A}}$ to be normal to the surface in either direction (such as for the surface bounded by the loop in Fig. 34-5). The solution to this dilemma is yet another right-hand rule: point the fingers of your right hand in the direction around the loop that you will use to calculate the emf. (For actual wire current loops, this is the same direction as the direction of the induced current.) Your thumb will then point in the direction that we take for $d\vec{\mathbf{A}}$ in calculating the flux (Fig. 34-7*b*).

With these definitions of the direction or sense of the emf and the direction of $d\vec{\mathbf{A}}$ for finding the flux, we can write Faraday's law as

$$\mathscr{E} = -\frac{d\Phi_B}{dt}. \tag{34-5}$$

The negative sign in this equation is in effect an expression of Lenz' law, so that the induced emf opposes the change in flux.

Figure 34-7*c* shows how to use Eq. 34-5. Suppose $\vec{\mathbf{B}}$ is downward through the loop and increasing in magnitude. Let us pick the upward direction for $d\vec{\mathbf{A}}$. (As long as we use the right-hand rule, it does not matter which direction we pick—it is the relationship between $\mathscr{E}$ and $d\mathbf{A}$ that is important.) For this choice $\vec{\mathbf{B}} \cdot d\vec{\mathbf{A}}$ is negative everywhere on the surface, so the flux is negative. If $\vec{\mathbf{B}}$ is increasing in magnitude, then Φ_B is negative and increasing in magnitude, so $d\Phi_B/dt < 0$. Equation 34-5 then tells us that $\mathscr{E}$ is positive relative to the direction we chose for $d\vec{\mathbf{A}}$. The right-hand rule shows that the emf (and thus the induced current) must be as indicated in Fig. 34-7*c*. The induced current would be counterclockwise as viewed from above, just as we would have concluded by applying Lenz' law directly.

SAMPLE PROBLEM 34-2. A circular loop of wire of radius $r = 0.32$ m and resistance $R = 2.5\ \Omega$ is placed in a uniform magnetic field that is perpendicular to the plane of the loop and initially directed into the page, as shown in Fig. 34-8. The field varies with time as $B_z = -4.0\ \text{T} - (5.6\ \text{T/s})t + (2.2\ \text{T/s}^2)t^2$, where the positive z axis is up out of the page. Find the magnitude and direction of the current in the loop at $t = 1$ s and at $t = 2$ s.

Solution Let us choose $d\vec{\mathbf{A}}$ into the page. At $t = 1$ s, $B_z = -7.4$ T, and at $t = 2$ s, $B_z = -6.4$ T. Thus for the time intervals covered by this problem, $\vec{\mathbf{B}}$ and $d\vec{\mathbf{A}}$ are parallel and $\Phi_B > 0$. Because the field is uniform over the loop, $\Phi_B = \vec{\mathbf{B}} \cdot \vec{\mathbf{A}} = B_z A_z$ (which is a positive quantity, because $A_z < 0$). Faraday's law then gives

$$\mathscr{E} = -\frac{d\Phi_B}{dt} = -\frac{d(B_z A_z)}{dt} = -A_z \frac{dB_z}{dt}.$$

The direction of the emf depends on the sign of $dB_z/dt = -5.6 + 4.4t$. At $t = 1$ s, $dB_z/dt = -1.2$ T/s. With $A = \pi r^2 = 0.322\ \text{m}^2$, we have $\mathscr{E} = -(-0.322\ \text{m}^2)(-1.2\ \text{T/m}) = -0.39$ V. The negative sign means that, with the thumb of the right hand pointed in the direction of $\vec{\mathbf{A}}$ (into the page), the direction of the emf (and thus the current) is counterclockwise. The magnitude of the current is $i = |\mathscr{E}|/R = 0.15$ A. You should check that the same direction for the current (counterclockwise) is obtained from Lenz' law.

Repeating the calculation for $t = 2$ s, we find $dB_z/dt = +3.2$ T/s and $\mathscr{E} = +1.03$ V. This gives a clockwise current of magnitude $i = |\mathscr{E}|/R = 0.37$ A.

Note that the direction of the current changes between $t = 1$ s and $t = 2$ s, even though the direction of the field remains the same (into the page). It is the direction of the *change* in the field rather than the direction of the field itself that determines the direction of the induced current.

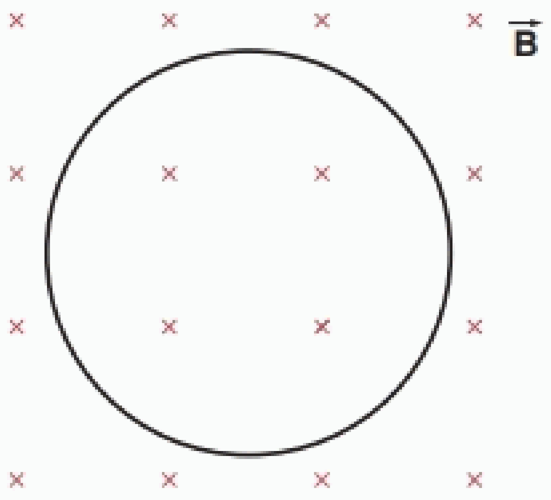


FIGURE 34-8. Sample Problem 34-2. A loop of wire in a magnetic field directed into the page.

34-4 MOTIONAL EMF

Figure 34-9 shows a rectangular loop of wire of width D, one end of which is in a uniform and constant field $\vec{\mathbf{B}}$ pointing at right angles to the plane of the loop. The field $\vec{\mathbf{B}}$ may be produced, for example, in the gap of a large electromagnet. The dashed lines show the assumed limits of the magnetic field. The loop is pulled to the right at a constant speed v.

The situation described by Fig. 34-9 does not differ in any essential detail from that of Fig. 34-5. In each case a conducting loop and a magnet are in relative motion; in each case the flux of the field of the magnet through the loop is being caused to change with time. The important difference between the two arrangements is that the situation of Fig. 34-9 permits easier calculations.

The external agent (the hand in Fig. 34-9) pulls the loop to the right at constant speed. As the loop moves, the portion of the area of the loop in the field decreases, and so the flux decreases. This decreasing flux induces an emf, and an induced current flows in the loop. This emf, which results from relative motion between a conductor and the source of the magnetic field, is sometimes called a *motional emf.*

The flux Φ_B enclosed by the loop in Fig. 34-9 is

$$\Phi_B = BA = BDx,$$

where Dx is the area of that part of the loop in which B is not zero. We find the emf $\mathcal{E}$ from Faraday's law:

$$|\mathcal{E}| = \left|\frac{d\Phi_B}{dt}\right| = \frac{d}{dt}(BDx) = BD\frac{dx}{dt} = BDv, \quad (34\text{-}6)$$

where we have set dx/dt equal to the speed v at which the loop is pulled out of the magnetic field. Note that the only dimension of the loop that enters into Eq. 34-6 is the length D of the left end conductor. As we shall see in Section 34-7, the induced emf in Fig. 34-9 may be regarded as localized in the left end of the loop.

The emf BDv sets up a current in the loop given by

$$i = \frac{|\mathcal{E}|}{R} = \frac{BDv}{R}, \quad (34\text{-}7)$$

where R is the loop resistance. From Lenz' law, this current must be clockwise in Fig. 34-9; it opposes the "change" (the decrease in Φ_B) by setting up a field that is parallel to the external field within the loop.

The sides of the loop, which we regard as current-carrying conductors, experience magnetic forces $\vec{\mathbf{F}}_B = i\vec{\mathbf{L}} \times \vec{\mathbf{B}}$ (see Eq. 32-26), as shown in Fig. 34-9. Because $\vec{\mathbf{F}}_2$ and $\vec{\mathbf{F}}_3$ are equal and opposite, their effects cancel. The net magnetic force on the loop is given by $\vec{\mathbf{F}}_1$, and in order to move the loop at constant speed the external agent must exert a force equal in magnitude and opposite in direction to $\vec{\mathbf{F}}_1$. Let us calculate the mechanical power $P = Fv$ that must be expended by the external agent or, equivalently, the rate at which the external agent does work on the loop.

The magnitude of the force $\vec{\mathbf{F}}_1$ is $F_1 = iDB$, because the direction of the current is always perpendicular to $\vec{\mathbf{B}}$. Since the external agent must exert a force equal to F_1, the power expended by the agent is

$$P = F_1 v = iDBv = \frac{B^2D^2v^2}{R}, \quad (34\text{-}8)$$

where we have used Eq. 34-7 for the induced current i.

We can also compute the rate at which energy is dissipated in the loop as a result of Joule heating by the induced current. This is given by

$$P = i^2R = \left(\frac{BDv}{R}\right)^2 R = \frac{B^2D^2v^2}{R}, \quad (34\text{-}9)$$

which agrees precisely with Eq. 34-8 for the rate at which mechanical work is done on the loop. The work done by the external agent is eventually dissipated as Joule heating of the loop.

Eddy Currents

When the magnetic flux through a large piece of conducting material changes, induced currents appear in the material (Fig. 34-10). These currents are called *eddy currents.* In

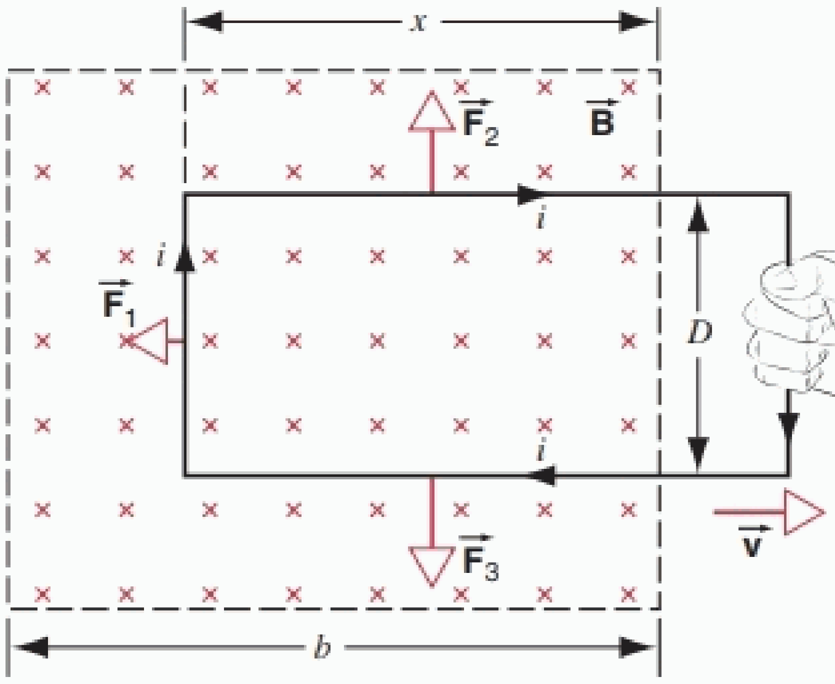


FIGURE 34-9. When the closed conducting loop is withdrawn from the field, an induced current i is produced in the loop.

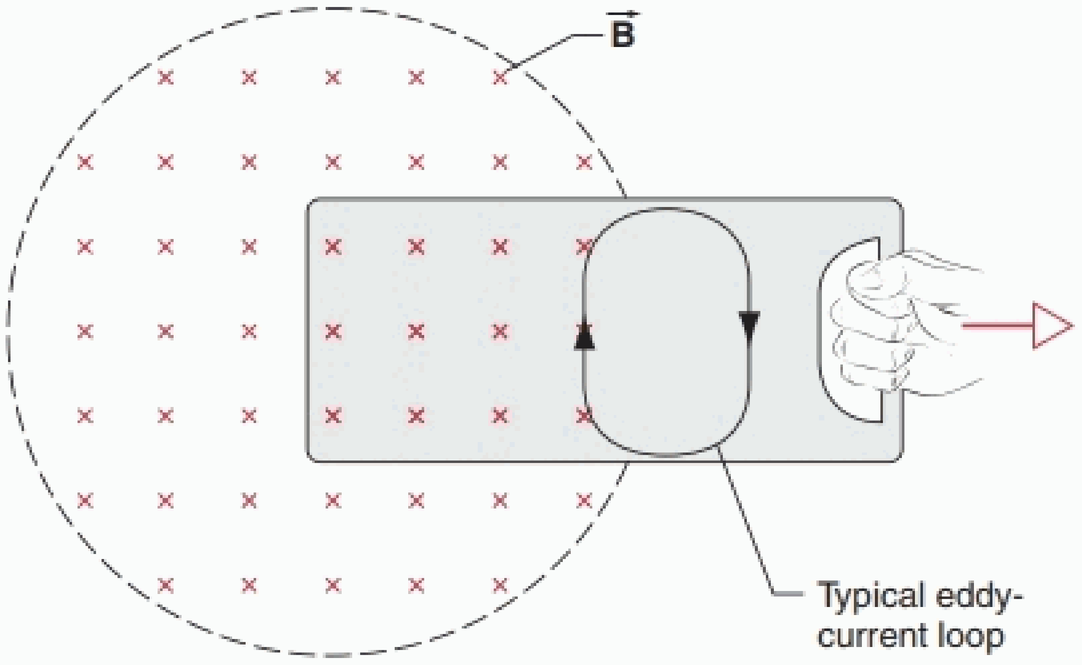


FIGURE 34-10. When the conducting material is withdrawn from the magnetic field, an induced current (eddy current) appears as shown.

some cases, the eddy currents may produce undesirable effects. For example, they increase the internal energy and thus can increase the temperature of the material. For this reason, materials that are subject to changing magnetic field are often *laminated* or constructed in many small layers insulated from one another. Instead of one large loop, the eddy currents follow many smaller loops, thereby increasing the total length of their paths and the corresponding resistance; the resistive heating $\mathcal{E}^2/R$ is smaller, and the increase in internal energy is smaller. On the other hand, eddy-current heating can be used to advantage, as in an *induction furnace,* in which a sample of material can be heated using a rapidly changing magnetic field. Induction furnaces are used in cases in which it is not possible to make thermal contact with the material to be heated, such as when it is enclosed in a vacuum chamber.

Eddy currents are real currents and produce the same effects as real currents. In particular, a force $\vec{\mathbf{F}}_B = i\vec{\mathbf{L}} \times \vec{\mathbf{B}}$ is exerted on the part of the eddy-current path in Fig. 34-10 that passes through the field. This force is transmitted to the material, and Lenz' law can be used to show (see Question 25) that the force opposes the motion of the conductor. This gives rise to a form of *magnetic braking,* in which magnetic fields applied to a rotating wheel or a moving track produce forces that decelerate the motion. Such a brake has no moving parts or mechanical linkages and is not subject to the frictional wear of ordinary mechanical brakes. Moreover, it is most efficient at high speed (because the magnetic force increases with the relative speed), where the wear on mechanical brakes would be greatest.

SAMPLE PROBLEM 34-3. Figure 34-11*a* shows a rectangular loop of resistance R, width D, and length a being pulled at constant speed v through a region of thickness d in which a uniform magnetic field $\vec{\mathbf{B}}$ is set up by a magnet. As functions of the position x of the right-hand edge of the loop, plot (*a*) the flux Φ_B through the loop, (*b*) the induced current i, and (*c*) the rate P of production of internal energy in the loop. Use $D = 4$ cm, $a = 10$ cm, $d = 15$ cm, $R = 16\ \Omega$, $B = 2.0$ T, and $v = 1.0$ m/s.

Solution (*a*) The flux Φ_B is zero when the loop is not in the field; it is BDa when the loop is entirely in the field; it is BDx when the loop is entering the field and $BD[a - (x - d)]$ when the loop is leaving the field. These conclusions, which you should verify, are shown graphically in Fig. 34-11*b*.

(*b*) The magnitude of the induced emf $\mathcal{E}$ is given by $\mathcal{E} = |d\Phi_B/dt|$, which we can write as

$$\mathcal{E} = \left|\frac{d\Phi_B}{dt}\right| = \left|\frac{d\Phi_B}{dx}\frac{dx}{dt}\right| = \left|\frac{d\Phi_B}{dx}v\right|,$$

where $d\Phi_B/dx$ is the slope of the curve of Fig. 34-11*b*. The current i is plotted as a function of x in Fig. 34-11*c*. Using the same type of reasoning as that used for Fig. 34-8, we deduce from Lenz' law that when the loop is entering the field, the current is counterclockwise as seen from above. Note that there is no current when the loop is entirely in the magnetic field because the flux Φ_B through the loop is not changing with time, as Fig. 34-11*b* shows.

(*c*) The rate of internal energy production is given by $P = i^2R$. It may be calculated by squaring the ordinate of the curve of Fig. 34-11*c* and multiplying by R. The result is plotted in Fig. 34-11*d*.

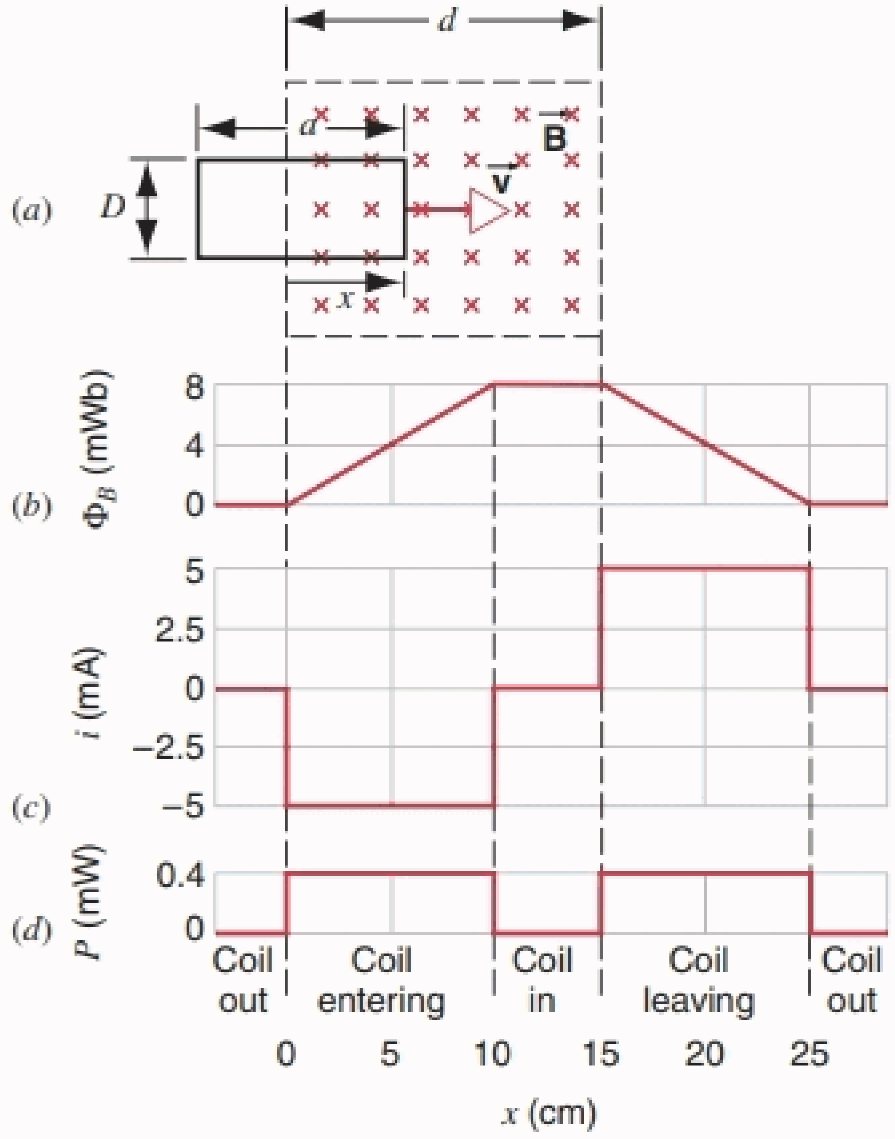


FIGURE 34-11. Sample Problem 34-3. (*a*) A closed conducting loop is pulled at constant speed completely through a region in which there is a uniform magnetic field $\vec{\mathbf{B}}$. (*b*) The magnetic flux through the loop as a function of the coordinate x of the right side of the loop. (*c*) The induced current as a function of x. Negative values indicate a counterclockwise current. (*d*) The rate at which internal energy appears in the loop as it is moved.

If the fringing of the magnetic field, which cannot be avoided in practice (see Problem 11 of Chapter 33), is taken into account, the sharp bends and corners in Fig. 34-11 will be replaced by smooth curves.

SAMPLE PROBLEM 34-4. A copper rod of length R rotates at angular frequency ω in a uniform magnetic field $\vec{\mathbf{B}}$ as shown in Fig. 34-12. Find the emf $\mathcal{E}$ developed between the two

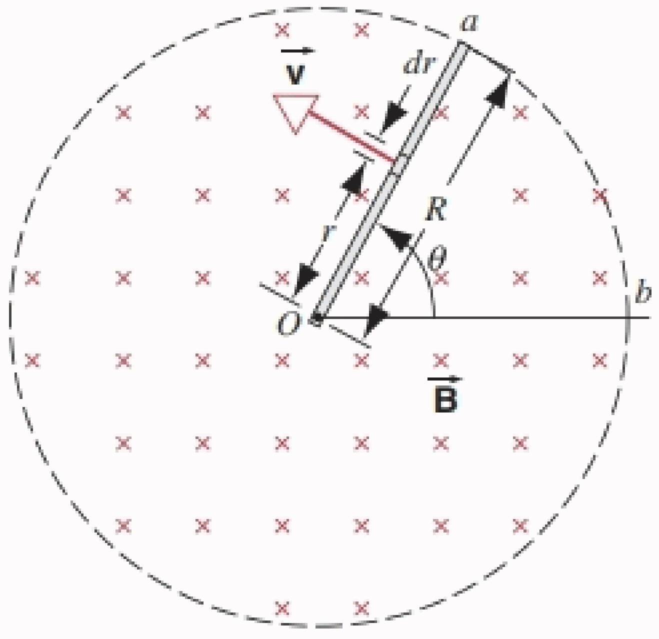


FIGURE 34-12. Sample Problem 34-4. A copper rod rotates in a uniform magnetic field.

ends of the rod. (We might measure this emf by placing a conducting rail along the dashed circle in the figure and connecting a voltmeter between the rail and point O.)

Solution If a wire of length dr is moved at velocity $\vec{\mathbf{v}}$ at right angles to a field $\vec{\mathbf{B}}$, a motional emf $d\mathcal{E}$ will be developed (see Eq. 34-6) given by

$$d\mathcal{E} = Bv\,dr.$$

The rod of Fig. 34-12 may be divided into elements of length dr, the linear speed v of each element being ωr. Each element is perpendicular to $\vec{\mathbf{B}}$ and is also moving in a direction at right angles to $\vec{\mathbf{B}}$ so that, since the emf's $d\mathcal{E}$ of each element are "in series,"

$$\mathcal{E} = \int d\mathcal{E} = \int_0^R Bv\,dr = \int_0^R B\omega r\,dr = \tfrac{1}{2}B\omega R^2.$$

For a second approach, consider that at any instant the flux enclosed by the sector aOb in Fig. 34-12 is given by

$$\Phi_B = BA = B(\tfrac{1}{2}R^2\theta),$$

where $\frac{1}{2}R^2\theta$ is the area of the sector. Differentiating gives

$$\frac{d\Phi_B}{dt} = \tfrac{1}{2}BR^2\frac{d\theta}{dt} = \tfrac{1}{2}B\omega R^2.$$

From Faraday's law, this is precisely the magnitude of $\mathcal{E}$ and agrees with the previous result.

34-5 GENERATORS AND MOTORS

As examples of the practical consequences of Faraday's law, we consider the operation of simple generators and motors.

Figure 34-13 shows the basic elements of a generator. A loop of conducting wire rotates with constant angular velocity ω in an external magnetic field. (Another device, not shown in the figure, is needed to cause the loop to rotate. In electrical power plants, the additional device might be falling water from a dam or steam produced in a boiler that turns the blades of a turbine.) For simplicity we assume that the magnetic field is uniform in the region in which the loop rotates.

The magnetic flux through the loop is given by Eq. 34-2: $\Phi_B = BA\cos\theta$. As the loop rotates, the angle θ between the directions of the magnetic field and the area elements $d\vec{\mathbf{A}}$ within the loop changes with time according to $\theta = \omega t$. The emf induced in the rotating loop is

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = -BA\frac{d}{dt}(\cos\omega t) = BA\omega\sin\omega t. \qquad (34\text{-}10)$$

If the loop has N turns, the total flux is multiplied by N, so that the emf would be $\mathcal{E} = NBA\omega\sin\omega t$.

The induced emf varies sinusoidally with time, as shown in Fig. 34-14. If the generator is connected to an external load of resistance R, an induced current $i = \mathcal{E}/R$ occurs in the circuit; this current flows through the rotating loop and the wires connected to the load.

Figure 34-14 suggests that the current changes direction as the loop rotates. A current that changes direction is called an *alternating current* (abbreviated AC). The emf produced by this generator is called an AC emf or an AC voltage.

Let us now consider the direction of the induced current in the loop. When the loop is in the position shown in Fig. 34-13, a small rotation in the direction of ω would decrease the flux, and so (by Lenz' law) the induced current in the loop must produce a field in the same direction as the external field (thereby opposing the decrease in the flux). The induced current is therefore counterclockwise, as shown. As the plane of the loop becomes parallel to the field ($\theta = 90°$) the flux is changing most rapidly and the counterclockwise current reaches its maximum magnitude. From $\theta = 90°$ to $\theta = 180°$ the flux is becoming increasingly negative, so the current must continue in the counterclockwise direction to oppose the change in flux. Eventually the loop rotates through $\theta = 180°$, and the negative flux begins to increase toward zero. The induced current must now be in the clockwise direction, producing a downward field inside the loop that opposes the change in flux. Continuing this reasoning, we conclude that the current changes direction whenever the loop rotates through 180°.

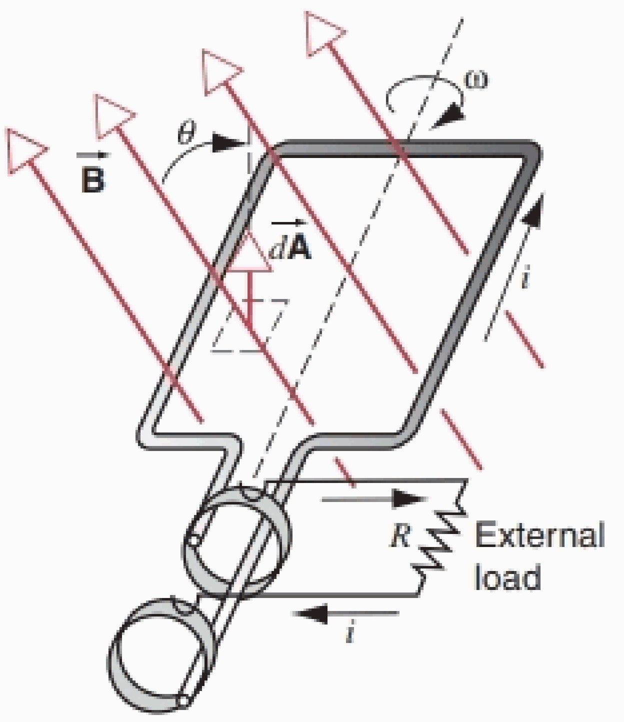


FIGURE 34-13. A simple generator. The rotation of the coil induces a current that alternates in direction. The current is delivered to an external load by sliding contacts (called "brushes") on the rings.

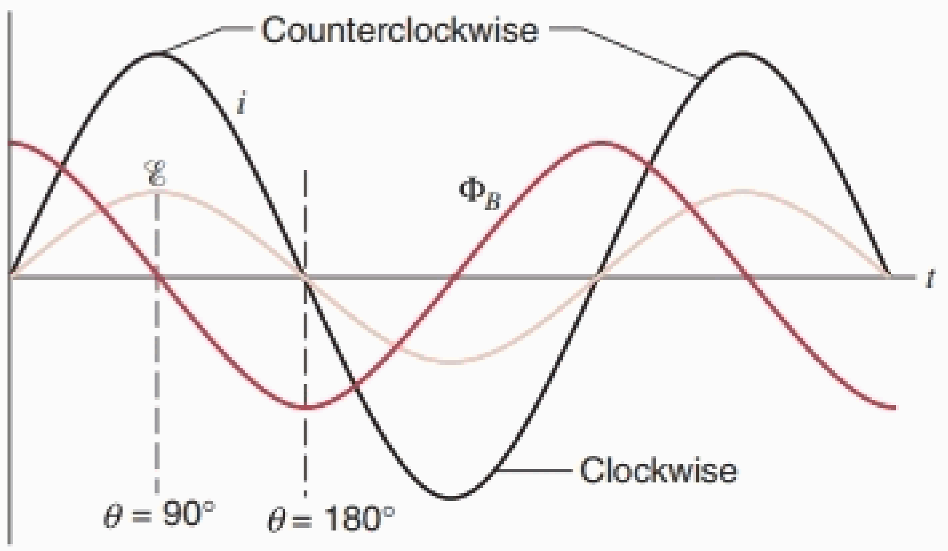


FIGURE 34-14. Magnetic flux, induced emf, and induced current for the generator of Fig. 34-13.

Generators of a more complex design than that shown here are used in electrical power plants to produce the AC voltage that we use in our homes. A similar device known as an *alternator* is used in automobiles; in an alternator the source of the magnetic field (a small permanent magnet) rotates, while the loop in which the current is induced remains fixed.

So far it appears that we can get "free" electricity from our generator. If we could design a rotating loop with frictionless bearings, once we set it rotating with a given angular velocity, the induced current should continue indefinitely in the external load. Unlimited work would be done in the external load with no investment of energy. Surely there must be something wrong with this reasoning, which violates conservation of energy!

Even in the absence of friction, there is a torque that inhibits the loop from rotating. When current flows in the loop, there is a magnetic torque given by Eq. 32-34 ($\tau = NiAB \sin \theta$). It does not matter whether the current is produced by a battery or by motion in a magnetic field; there is still a torque on the current loop. In the situation shown in Fig. 34-13, the torque acts to push the plane of the loop toward $\theta = 0$, and so it opposes the rotation. In fact, you should convince yourself that the torque remains in the same direction even when the loop passes through $\theta = 180°$ and the current changes direction. To oppose this torque, the device that is producing the rotation of the loop must continue to do work as the loop rotates. The generator can thus be viewed as a device for converting mechanical work (the rotation of the loop) into electrical work in the load. The electrical energy produced by the generator is ultimately derived from the energy supplied by the agent that keeps the loop rotating.

An electric motor is simply a generator operating in reverse. We disconnect the external source that is driving the rotation of the loop, and we replace the load in Fig. 34-13 with another generator, which produces an AC emf $\mathcal{E}$ that drives a current $i = \mathcal{E}/R$ in the loop. In this case there is again a magnetic torque on the loop that causes it to rotate. Just as the loop passes through $\theta = 180°$, where the torque is zero, the externally provided current changes direction, which keeps the torque in the same direction as the loop continues to rotate. Even though the current changes direction every half cycle, the magnetic torque remains in the same direction.

Once again, if we assume frictionless bearings, it appears we should be able to get something for nothing from our motor. If we make the resistance of the wires very small, then the current and the resulting torque become very large. An unlimited amount of mechanical work appears to be possible with only a small investment of electrical energy. What we have neglected is that the rotating loop produces an induced emf $\mathcal{E}_{\text{ind}}$ (known in the case of a motor as a "back emf") given by Eq. 34-10. According to Lenz' law, the back emf opposes the effect of the applied emf $\mathcal{E}$. When the motor first starts, ω is small, the back emf is small, and the current is $i = \mathcal{E}/R$. As the rotation picks up speed, the back emf increases and the current decreases to $i = (\mathcal{E} - \mathcal{E}_{\text{ind}})/R$. As the rotation speed continues to increase, the back emf increases; when eventually $\mathcal{E}_{\text{ind}} = \mathcal{E}$, no current flows and the rotating motor can no longer supply any torque. If we apply a load to the motor (for example, a weight to be lifted), the rotation slows somewhat, so $\mathcal{E}_{\text{ind}}$ decreases and i increases—the source generator must provide additional electrical work. A motor can therefore be regarded as a device for converting electrical work (from the driving generator) into mechanical work.

Real generators and motors are somewhat more complicated than the ones we have discussed. Some generators use clever geometric arrangements of coil windings and pickup mechanisms to produce DC currents (which may vary in magnitude with time but do not change direction). Similarly, there are DC motors that operate with DC currents or voltages. Nevertheless, the basic operating principles are similar to the examples discussed here.

SAMPLE PROBLEM 34-5. An electrical generator consists of a rectangular loop of dimensions 8.4 cm by 15.4 cm. It rotates in a uniform magnetic field of 0.126 T at a frequency of 60.0 Hz about an axis perpendicular to the field direction. What is the maximum emf generated by the loop?

Solution The maximum emf is given by Eq. 34-10 when $\sin \omega t = 1$:

$$\begin{aligned}\mathcal{E}_{\max} &= BA\omega \\ &= (0.126\ \text{T})(0.084\ \text{m} \times 0.154\ \text{m})(2\pi \times 60\ \text{s}^{-1}) \\ &= 0.61\ \text{V}.\end{aligned}$$

34-6 INDUCED ELECTRIC FIELDS

Suppose we place a loop of conducting wire in an external magnetic field (as in Fig. 34-15*a*). The field, which we assume to have a uniform strength over the area of the loop, may be established by an external electromagnet. By varying the current in the electromagnet, we can vary the strength of the magnetic field.

As $\vec{\mathbf{B}}$ is varied, the magnetic flux through the loop varies with time, and from Faraday's and Lenz' laws we can calculate the magnitude and direction of the induced emf and the induced current in the loop. Before the field began changing, there was no current in the loop; while the field is changing, charges flow in the loop. For charges to begin moving, they must be accelerated by an electric field. This *induced electric field* occurs with a changing magnetic field, according to Faraday's law.

The induced electric field is as real as any that might be set up by static charges; for instance, it exerts a force $q_0\vec{\mathbf{E}}$ on a test charge. Moreover, the presence of the electric field has nothing to do with the presence of the loop of wire; if

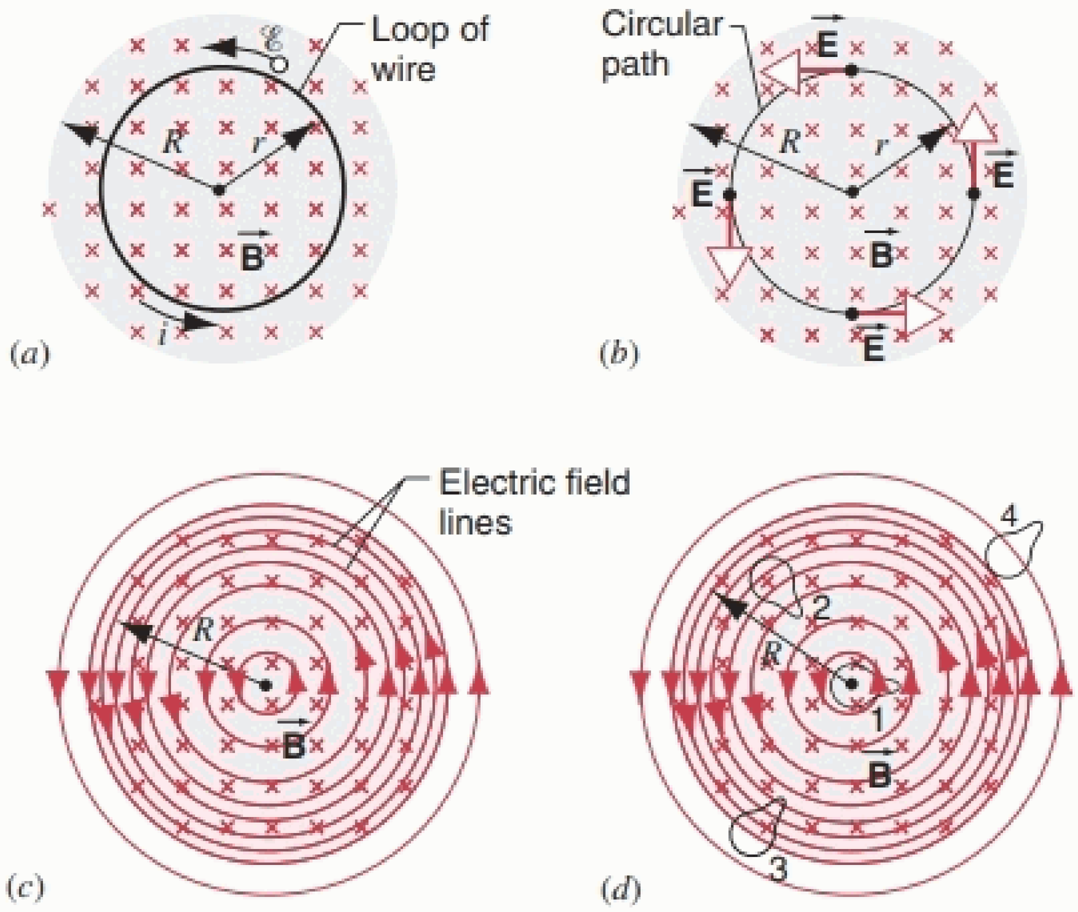


FIGURE 34-15. (*a*) If the magnetic field (which points into the page) increases at a steady rate, a constant current appears, as shown, in the loop of wire of radius r. (*b*) Induced electric fields exist in the region, even when the ring is removed. (*c*) The complete picture of the induced electric fields, displayed as field lines. (*d*) Four similar closed paths around which an emf may be calculated.

we were to remove the loop completely, the electric field would still be present. We could fill the space with a "gas" of electrons or ionized atoms; these particles would experience this induced electric field $\vec{\mathbf{E}}$.

Let us therefore replace the loop of wire with a circular path of arbitrary radius r (Fig. 34-15*b*). The path, which we take in a plane perpendicular to the direction of $\vec{\mathbf{B}}$, encloses a region of space in which the magnetic field is changing at a rate $d\vec{\mathbf{B}}/dt$. We assume that the rate $d\vec{\mathbf{B}}/dt$ is the same at every point in the area enclosed by the path. The circular path encloses a flux Φ_B, which is changing at a rate $d\Phi_B/dt$ because of the variation in the magnetic field. An induced emf appears around the path, and therefore there is an induced electric field at all points around the circle. From symmetry, we conclude that $\vec{\mathbf{E}}$ must have the same magnitude at all points around the circle, there being no preferred direction in this space. Furthermore, $\vec{\mathbf{E}}$ can have no radial component, a conclusion that follows from Gauss' law: construct an imaginary cylindrical Gaussian surface perpendicular to the plane of Fig. 34-15*b*. If there were a radial component to $\vec{\mathbf{E}}$, there would be a net *electric* flux into or out of the surface, which would require that the surface enclose a net electric charge. Since there is no such charge, the electric flux must be zero and the radial component of $\vec{\mathbf{E}}$ must be zero. Thus the induced electric field is tangential, and the electric field lines are concentric circles, as in Fig. 34-15*c*.

Consider a test charge q_0 moving around the circular path of Fig. 34-15*b*. The work W done on the charge by the induced electric field in one revolution is $\mathcal{E}q_0$. Equivalently, we can express the work as the electric force q_0E times the displacement $2\pi r$ covered in one revolution. Setting these two expressions for W equal to one another and canceling the factor q_0, we obtain

$$\mathcal{E} = E(2\pi r). \tag{34-11}$$

The right side of Eq. 34-11 can be expressed as a line integral of $\vec{\mathbf{E}}$ around the circle, which can be written in more general cases (for instance, when $\vec{\mathbf{E}}$ is not constant or when the chosen closed path is not a circle) as

$$\mathcal{E} = \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}. \tag{34-12}$$

Note that Eq. 34-12 reduces directly to Eq. 34-11 in our special case of a circular path with constant tangential $\vec{\mathbf{E}}$.

Replacing the emf by Eq. 34-12, we can write Faraday's law of induction ($\mathcal{E} = -d\Phi_B/dt$) as

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -\frac{d\Phi_B}{dt}. \tag{34-13}$$

It is in this form that Faraday's law appears as one of the four basic Maxwell equations of electromagnetism. In this form, it is apparent that Faraday's law implies that a changing magnetic field produces an electric field. The direction of the line integral is related to the direction of $d\vec{\mathbf{A}}$ in Φ_B by the same right-hand rule discussed at the end of Section 34-3: with the fingers in the direction of $d\vec{\mathbf{s}}$ around the closed integration path, the thumb indicates the direction of $d\vec{\mathbf{A}}$.

In Fig. 34-15, we have assumed that the magnetic field is increasing; that is, both dB/dt and $d\Phi_B/dt$ are positive. By Lenz' law, the induced emf opposes this change, and thus the induced currents create a magnetic field that points out of the plane of the figure. Since the currents must be counterclockwise, the lines of induced electric field $\vec{\mathbf{E}}$ (which is responsible for the current) must also be counterclockwise. If, on the other hand, the magnetic field were decreasing ($dB/dt < 0$), the lines of induced electric field would be clockwise, such that the induced current again opposes the change in Φ_B.

Faraday's law in the form of Eq. 34-13 can be applied to paths of any geometry, not only the special circular path we chose in Fig. 34-15*b*. Figure 34-15*d* shows four such paths, all having the same shape and area but located in different positions in the changing field. For paths 1 and 2, the induced emf is the same because these paths lie entirely within the changing magnetic field and thus have the same value of $d\Phi_B/dt$. However, even though the emf $\mathcal{E}$ $(=\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}})$ is the same for these two paths, the distribution of electric field vectors around the paths is different, as indicated by the lines of the electric field. For path 3, the emf is smaller because both Φ_B and $d\Phi_B/dt$ are smaller, and for path 4 the induced emf is zero, even though the electric field is not zero at any point along the path.

Induced EMF and Potential Difference

In Chapter 26 we discussed electric fields produced by charges. Induced electric fields are produced not by charges but by a changing magnetic flux. Both kinds of electric fields can be detected by the forces they exert on charges, but there is an important difference between them: the lines of $\vec{\mathbf{E}}$ associated with a changing magnetic flux form closed loops, but the lines of $\vec{\mathbf{E}}$ associated with charges always start on a positive charge and end on a negative charge.

There is another difference between the two kinds of electric fields: *electric fields produced by charges can be represented by a potential, but potential has no meaning for electric fields produced by a changing magnetic flux.* In Chapter 28 we showed that the potential difference between two points in an electric field is (see Eq. 28-15):

$$V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}. \qquad (34\text{-}14)$$

Because the electrostatic field associated with charges is conservative, the potential difference is independent of the path we travel between points a and b. If a and b are the same point, then the path connecting them is a closed loop, and Eq. 34-14 becomes:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = 0. \qquad (34\text{-}15)$$

Figure 34-16a illustrates this concept. If we connect the two leads of a voltmeter together in a field due to charges, the voltmeter reads zero.

Consider the contrasting case of the electric field produced by a changing magnetic field. In this case the integral of $\vec{\mathbf{E}}$ around the closed path is *not* zero—now $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}$ is $-d\Phi_B/dt$, according to Faraday's law. In Fig. 34-16b, the two leads of the voltmeter, still connected together, form a loop surrounding a solenoid in which the current is changing. In this case the voltmeter reading is *not* zero, and the concept of potential cannot be used to describe the situation. The induced electric field due to a changing magnetic field is nonconservative and cannot be represented by a potential. (The magnetic field due to a current is also nonconservative. The lines of magnetic field form closed loops, and the magnetic field also cannot be described by a potential.)

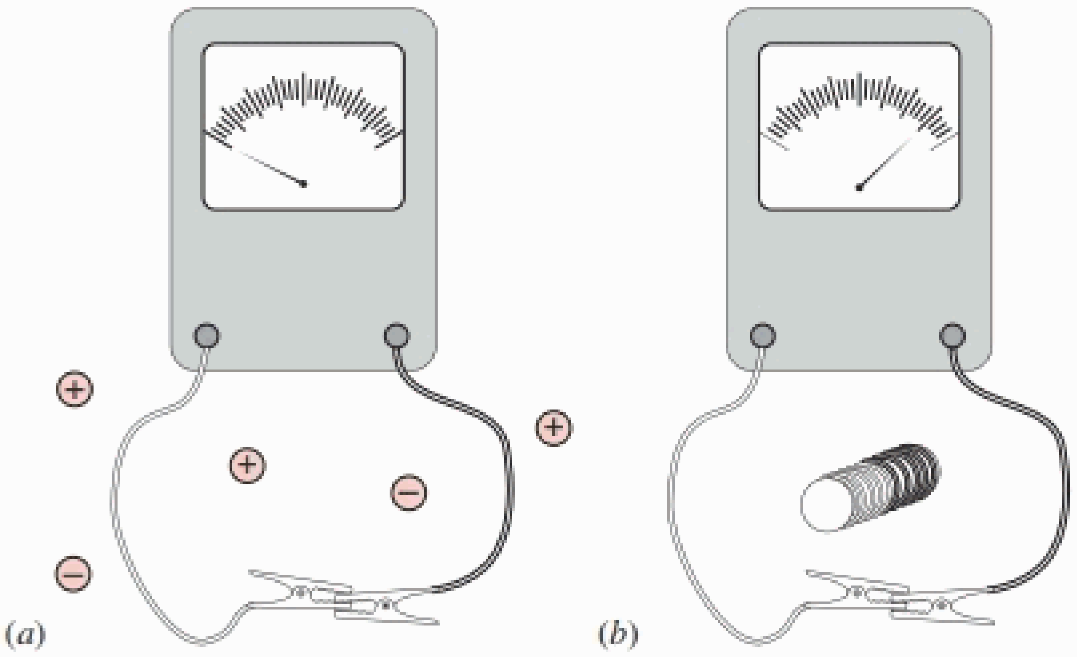


FIGURE 34-16. (a) The leads of a voltmeter are connected together in a region where there is an electric field due to charges. The integral $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}$ around the closed path consisting of the leads and the voltmeter gives zero, so the voltmeter reads zero. (b) The leads now surround a solenoid in which the current is changing. Even though the leads are connected together, the voltmeter reading is not zero.

SAMPLE PROBLEM 34-6. In Fig. 34-15b, assume that $R = 8.5$ cm and that $dB/dt = 0.13$ T/s. (a) What is the magnitude of the electric field $\vec{\mathbf{E}}$ for $r = 5.2$ cm? (b) What is the magnitude of the induced electric field for $r = 12.5$ cm?

Solution (a) The flux Φ_B through a closed circular path of radius r is (with $r < R$)

$$\Phi_B = B(\pi r^2),$$

so that Faraday's law (Eq. 34-13) gives

$$E(2\pi r) = -\frac{d\Phi_B}{dt} = -(\pi r^2)\frac{dB}{dt}.$$

Solving for E and taking magnitudes, we find

$$E = \frac{1}{2}\left|\frac{dB}{dt}\right| r. \qquad (34\text{-}16)$$

Note that the induced electric field E depends on dB/dt but not on B. For $r = 5.2$ cm, we have, for the magnitude of $\vec{\mathbf{E}}$,

$$E = \frac{1}{2}\left|\frac{dB}{dt}\right| r = \tfrac{1}{2}(0.13\ \text{T/s})(5.2 \times 10^{-2}\ \text{m})$$

$$= 0.0034\ \text{V/m} = 3.4\ \text{mV/m}.$$

(b) In this case we have $r > R$ so that the entire flux of the magnet passes through the circular path. Thus

$$\Phi_B = B(\pi R^2).$$

From Faraday's law (Eq. 34-13) we then find

$$E(2\pi r) = -\frac{d\Phi_B}{dt} = -(\pi R^2)\frac{dB}{dt}.$$

Solving for E and again taking magnitudes, we find

$$E = \frac{1}{2}\left|\frac{dB}{dt}\right| \frac{R^2}{r}. \qquad (34\text{-}17)$$

An electric field is induced in this case even at points that are well outside the (changing) magnetic field, an important result that makes transformers (see Section 37-5) possible. For $r = 12.5$ cm, Eq. 34-17 gives

$$E = \tfrac{1}{2}(0.13\ \text{T/s})\,\frac{(8.5 \times 10^{-2}\ \text{m})^2}{12.5 \times 10^{-2}\ \text{m}}$$

$$= 3.8 \times 10^{-3}\ \text{V/m} = 3.8\ \text{mV/m}.$$

Equations 34-16 and 34-17 yield the same result, as they must, for $r = R$. Figure 34-17 shows a plot of $E(r)$ based on these two equations.

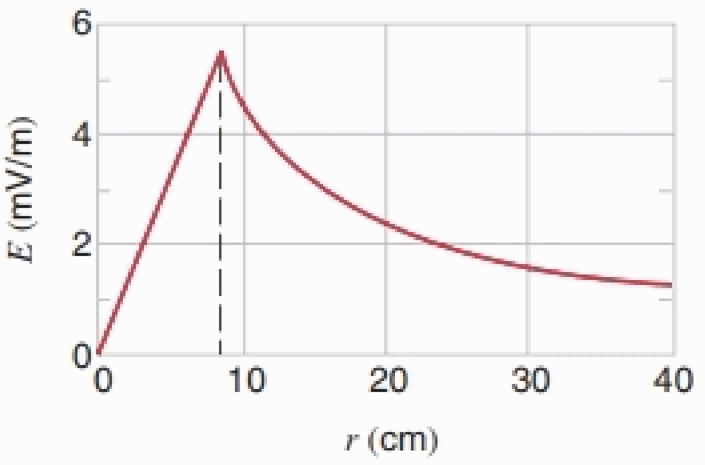


FIGURE 34-17. The induced electric field found in Sample Problem 34-6.

The Betatron*

The betatron is a device for accelerating electrons (also known as beta particles) to high speeds using the induced electric field produced by a changing magnetic field. Typical electron energies are 50–100 MeV. Such high-energy electrons can be used for basic research in physics as well as for producing x rays for applied research in industry and for medical purposes such as cancer therapy.

Figure 34-18*a* shows a cross section through the inner structure of a betatron. The magnetic field, whose shape is determined by the pole pieces M made of magnetic material, can be varied by changing the current in the coils. The electrons circulate in the evacuated, doughnut-shaped tube.

The coils carry an alternating current and produce the magnetic field shown in Fig. 34-18*b*. For electrons to circulate in the direction shown (counterclockwise as viewed from above), the magnetic field must point upward (which we take as positive). The changing field must have $dB/dt > 0$, so that $d\Phi_B/dt > 0$ and the electrons are accelerated (rather than decelerated) during the cycle. Thus only the first quarter-cycle of Fig. 34-18*b* is useful for the operation of the betatron. Electrons enter the machine at $t = 0$ and are extracted at $t = T/4$; for the remaining three-quarters of the cycle, the betatron produces no beam. The betatron produces a pulsed, rather than a continuous beam, with pulses typically every 0.01 s.

SAMPLE PROBLEM 34-7. In a 100-MeV betatron, the orbit radius R is 84 cm. The magnetic field in the region enclosed by the orbit rises periodically (60 times per second) from zero to a maximum average value $B_{\text{av,m}} = 0.80$ T in an accelerating interval of one-fourth of a period, or 4.2 ms. (*a*) How much energy does the electron gain in one average trip around its orbit in this changing flux? (*b*) What is the *average* speed of an electron during its acceleration cycle?

Solution (*a*) The central flux rises during the accelerating interval from zero to a maximum of

$$\Phi_B = (B_{\text{av,m}})(\pi R^2)$$
$$= (0.80\ \text{T})(\pi)(0.84\ \text{m})^2 = 1.8\ \text{Wb}.$$

The average value of $d\Phi_B/dt$ during the accelerating interval is then

$$\left(\frac{d\Phi_B}{dt}\right)_{\text{av}} = \frac{\Delta\Phi_B}{\Delta t} = \frac{1.8\ \text{Wb}}{4.2\times 10^{-3}\ \text{s}} = 430\ \text{Wb/s}.$$

From Faraday's law (Eq. 34-3) this is also the average emf in volts. Thus the electron increases its energy by an average of 430 eV per revolution in this changing flux. To achieve its full final energy of 100 MeV, it has to make about 230,000 revolutions in its orbit, a total path length of about 1200 km.

(*b*) The length of the acceleration cycle is given as 4.2 ms, and the path length is calculated above to be 1200 km. The average speed is then

$$v_{\text{av}} = \frac{1200\times 10^3\ \text{m}}{4.2\times 10^{-3}\ \text{s}} = 2.86\times 10^8\ \text{m/s}.$$

This is 95% of the speed of light. The actual speed of the fully accelerated electron, when it has reached its final energy of 100 MeV, is 99.9987% of the speed of light.

34-7 INDUCTION AND RELATIVE MOTION (Optional)

In Section 33-6, we discussed the idea that the classification of electromagnetic effects into purely electric or purely magnetic was dependent on the reference frame of the observer. For example, what appears to be a magnetic field in

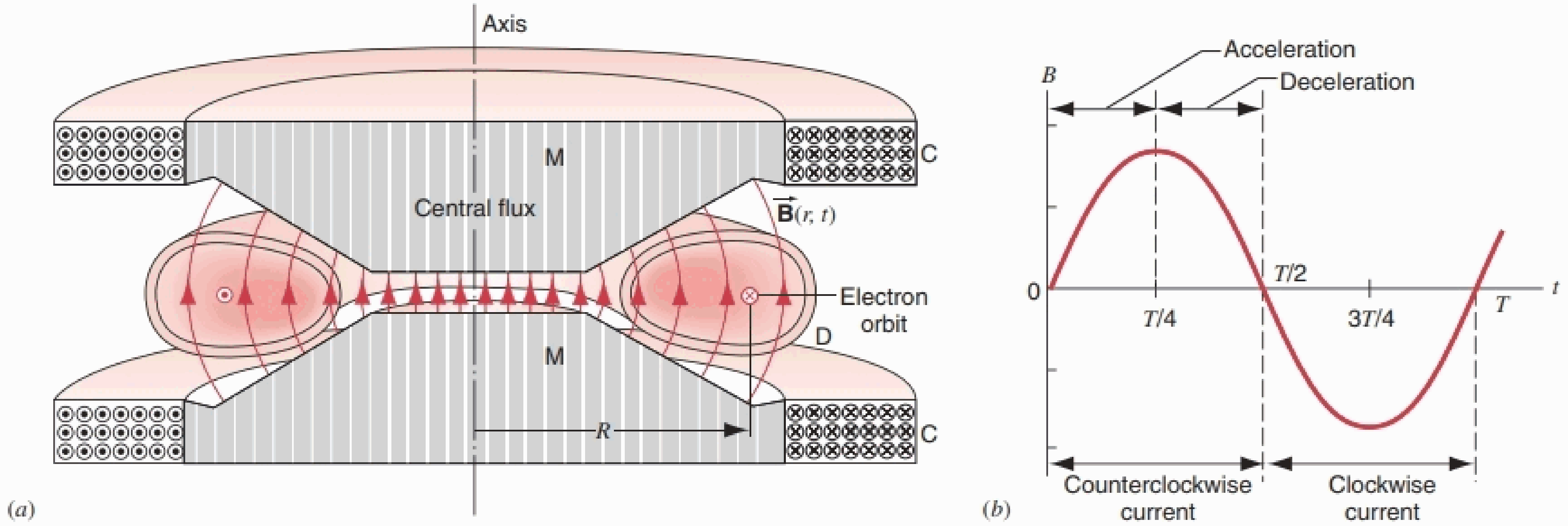


FIGURE 34-18. (*a*) Cross section of a betatron, showing the orbit of the accelerating electrons and a "snapshot" of the magnetic field at a certain time during the cycle. The magnetic field is produced by the coils C and shaped by the magnetic pole pieces M. Electrons circulate within the evacuated, doughnut-shaped ceramic tube D. Electrons orbit perpendicular to the plane of the figure, entering at right and leaving at left. (*b*) The variation with time of the betatron magnetic field during one cycle.

---

*For a review of developments and applications of betatrons and similar devices, see "Ultra-high-current Electron Induction Accelerators," by Chris A. Kapetanakos and Phillip Sprangle, *Physics Today,* February 1985, p. 58.

one frame of reference can appear as a mixture of electric and magnetic fields in another frame of reference. Since motional emf is determined by the velocity of the object moving through the magnetic field, it clearly depends on the reference frame of the observer. Other observers in different inertial frames will measure different velocities and different magnetic field strengths. It is therefore essential in calculating induced emfs and currents to specify the reference frame of the observer.

Figure 34-19*a* shows a closed loop, which an external agent (not shown) causes to move at velocity $\vec{\mathbf{v}}$ with respect to a magnet that provides a uniform field $\vec{\mathbf{B}}$ over a region. An observer S is at rest with respect to the magnet used to establish the field $\vec{\mathbf{B}}$. The induced emf in this case is a *motional emf* because the conducting loop is moving with respect to this observer.

Consider a positive charge carrier in the left end of the loop. To observer S, this charge q is constrained to move through the field $\vec{\mathbf{B}}$ with velocity $\vec{\mathbf{v}}$ to the right along with the loop, and it experiences a magnetic force given by $\vec{\mathbf{F}} = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}$ (not shown in Fig. 34-19*a*). This force causes the carriers to move upward (in the y direction) along the conductor; eventually, they acquire the drift velocity $\vec{\mathbf{v}}_{\mathrm{d}}$, as shown in Fig. 34-19*a*.

The resultant equilibrium velocity of the carriers is now $\vec{\mathbf{V}}$, the vector sum of $\vec{\mathbf{v}}$ and $\vec{\mathbf{v}}_{\mathrm{d}}$. In this situation the magnetic force $\vec{\mathbf{F}}_B$ is

$$\vec{\mathbf{F}}_B = q\vec{\mathbf{V}} \times \vec{\mathbf{B}}, \tag{34-18}$$

which is acting (as usual) at right angles to the resultant velocity $\vec{\mathbf{V}}$ of the carrier, as shown in Fig. 34-19*a*.

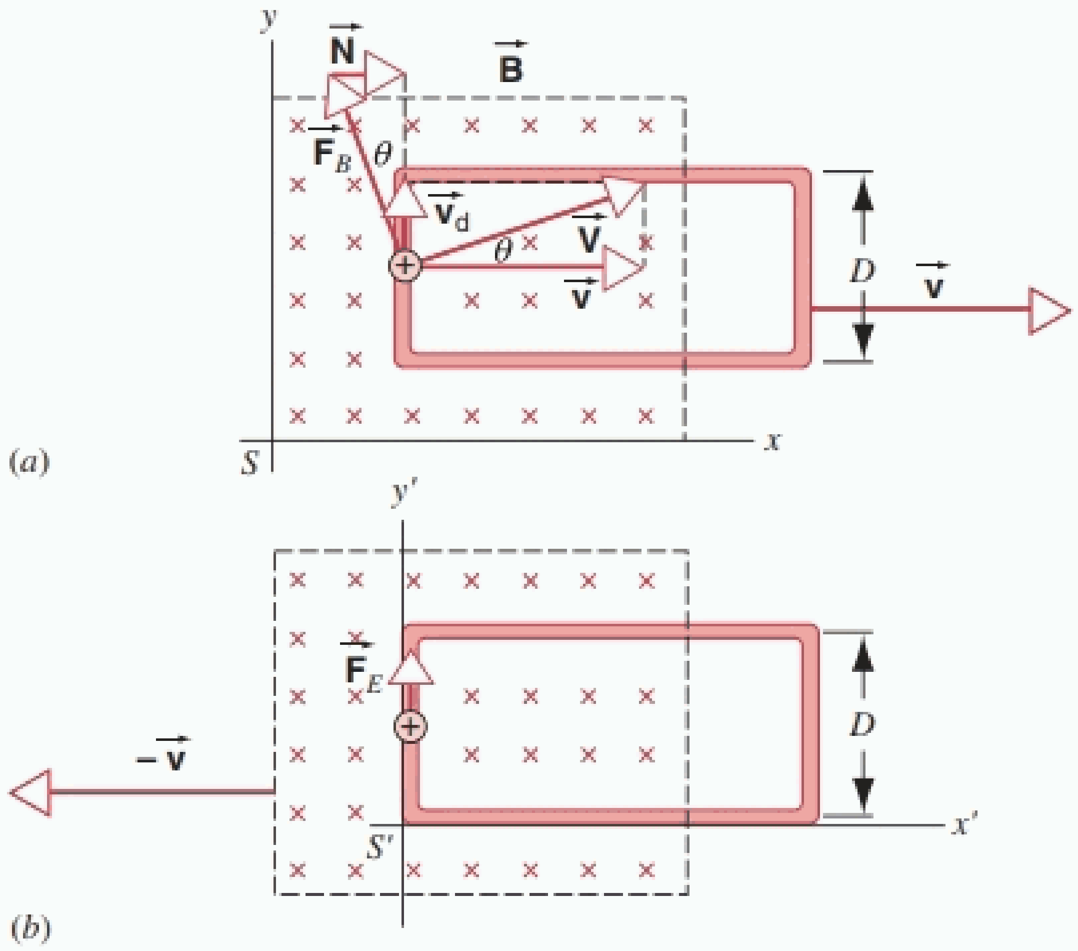


FIGURE 34-19. A closed conducting loop is in motion with respect to a magnet that produces the field $\vec{\mathbf{B}}$. (*a*) An observer S, fixed with respect to the magnet, sees the loop moving to the right and observes a magnetic force $F_B \cos\theta$ acting upward on the positive charge carriers. (*b*) An observer S', fixed with respect to the loop, sees the magnet moving toward the left and observes an *electric* force acting upward on the positive charge carriers. In both figures there are internal forces of collision (not shown) that keep the charge carriers from accelerating.

Acting alone, $\vec{\mathbf{F}}_B$ would tend to push the carriers through the left wall of the conductor. Because this does not happen, the conductor wall must exert a normal force $\vec{\mathbf{N}}$ on the carriers (see Fig. 34-19*a*) of magnitude such that $\vec{\mathbf{v}}_{\mathrm{d}}$ lies parallel to the axis of the wire; in other words, $\vec{\mathbf{N}}$ exactly cancels the horizontal component of $\vec{\mathbf{F}}_B$, leaving only the component $F_B \cos\theta$ that lies along the direction of the conductor. This latter component of force on the carrier is also canceled out in this case by the average impulsive force F_{i} associated with the internal collisions that the carrier experiences as it drifts with (constant) speed v_{d} through the wire.

The kinetic energy of the charge carrier as it drifts through the wire remains constant. This is consistent with the fact that the resultant force acting on the charge carrier $(= \vec{\mathbf{F}}_B + \vec{\mathbf{F}}_{\mathrm{i}} + \vec{\mathbf{N}})$ is zero. The work done by $\vec{\mathbf{F}}_B$ is zero because magnetic forces, acting at right angles to the velocity of a moving charge, can do no work on that charge. Thus the (negative) work done on the carrier by the average internal collision force $\vec{\mathbf{F}}_{\mathrm{i}}$ must be exactly canceled by the (positive) work done on the carrier by the force $\vec{\mathbf{N}}$. Ultimately, $\vec{\mathbf{N}}$ is supplied by the agent that pulls the loop through the magnetic field, and the mechanical energy expended by this agent appears as internal energy in the loop, as we have seen in Section 34-4.

Let us then calculate the work dW done on the carrier in time dt by the force $\vec{\mathbf{N}}$; it is

$$dW = N(v\,dt), \tag{34-19}$$

in which $v\,dt$ is the distance that the loop (and the carrier) has moved to the right in Fig. 34-19*a* in time dt. We can write for N (see Eq. 34-18 and Fig. 34-19*a*)

$$N = F_B \sin\theta = (qVB)(v_{\mathrm{d}}/V) = qBv_{\mathrm{d}}. \tag{34-20}$$

Substituting Eq. 34-20 into Eq. 34-19 yields

$$\begin{aligned} dW &= (qBv_{\mathrm{d}})(v\,dt) \\ &= (qBv)(v_{\mathrm{d}}\,dt) = qBv\,ds, \end{aligned} \tag{34-21}$$

in which $ds\ (= v_{\mathrm{d}}\,dt)$ is the distance the carrier drifts along the conductor in time dt.

The work done on the carrier as it makes a complete circuit of the loop is found by integrating Eq. 34-21 around the loop and is

$$W = \oint dW = qBvD. \tag{34-22}$$

This follows because work contributions for the top and the bottom of the loop are opposite in sign and cancel, and no work is done in those portions of the loop that lie outside the magnetic field.

An agent that does work on charge carriers, thus establishing a current in a closed conducting loop, can be viewed as an emf. Using Eq. 34-22, we find

$$\mathcal{E} = \frac{W}{q} = \frac{qBvD}{q} = BDv, \tag{34-23}$$

which is the same result that we derived from Faraday's law of induction; see Eq. 34-6. Thus a motional emf is intimately connected with the sideways deflection of a charged particle moving through a magnetic field.

We now consider how the situation of Fig. 34-19*a* would appear to an observer S' who is *at rest with respect to the loop*. To this observer, the magnet is moving to the left in Fig. 34-19*b* with velocity $-\vec{\mathbf{v}}$, and the charge q does not move in the x' direction with the loop but drifts clockwise around the loop. S' measures an emf $\mathscr{E}'$ that is accounted for, at the microscopic level, by postulating that an electric field $\vec{\mathbf{E}}'$ is induced in the loop by the action of the moving magnet. The emf $\mathscr{E}'$ is related to $\vec{\mathbf{E}}'$ by Eq. 34-12,

$$\mathscr{E}' = \oint \vec{\mathbf{E}}' \cdot d\vec{\mathbf{s}}. \tag{34-24}$$

The induced field $\vec{\mathbf{E}}'$, which has the same origin as the induced fields that we discussed in Section 34-6, exerts a force $q\vec{\mathbf{E}}'$ on the charge carrier.

The induced field $\vec{\mathbf{E}}'$ that produces the current exists only in the left end of the loop. (As we carry out the integral of Eq. 34-12 around the loop, contributions to the integral from the x' component of $\vec{\mathbf{E}}'$ cancel on the top and bottom sides, whereas there is no contribution from parts of the loop that are not in the magnetic field.) Using Eq. 34-12 we then obtain

$$\mathscr{E}' = E'D. \tag{34-25}$$

For motion at speeds that are small compared with the speed of light, the emfs given by Eqs. 34-23 and 34-25 must be identical, because the relative motion of the loop and the magnet is identical in the two cases shown in Fig. 34-19. Equating these relations yields

$$E'D = BDv,$$

or

$$E' = vB. \tag{34-26}$$

In Fig. 34-19*b* the vector $\vec{\mathbf{E}}'$ points upward along the axis of the left end of the conducting loop because this is the direction in which positive charges are observed to drift. The directions of $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$ are clearly shown in this figure. We see then that Eq. 34-26 is consistent with the more general vector relation

$$\vec{\mathbf{E}}' = \vec{\mathbf{v}} \times \vec{\mathbf{B}}. \tag{34-27}$$

We have not proved Eq. 34-27 except for the special case of Fig. 34-19; nevertheless it is true in general, no matter what the angle between $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$.

We interpret Eq. 34-27 in the following way. Observer S fixed with respect to the magnet is aware only of a magnetic field. To this observer, the force arises from the motion of the charges through $\vec{\mathbf{B}}$. Observer S' fixed on the charge carrier is aware of an electric field $\vec{\mathbf{E}}'$ also and attributes the force on the charge (at rest initially with respect to S') to the electric field. S says the force is of purely magnetic origin, whereas S' says the force is of purely electric origin. From the point of view of S, the induced emf is given by $\oint (\vec{\mathbf{v}} \times \vec{\mathbf{B}}) \cdot d\vec{\mathbf{s}}$. From the point of view of S', the same induced emf is given by $\oint \vec{\mathbf{E}}' \cdot d\vec{\mathbf{s}}$, where $\vec{\mathbf{E}}'$ is the (induced) electric field vector that S' observes at points along the circuit.

For a third observer S'', relative to whom both the magnet and the loop are moving, the force tending to move charges around the loop is neither purely electric nor purely magnetic, but a bit of each. In summary, in the equation

$$\vec{\mathbf{F}}/q = \vec{\mathbf{E}} + \vec{\mathbf{v}} \times \vec{\mathbf{B}},$$

different observers form different assessments of $\vec{\mathbf{E}}$, $\vec{\mathbf{B}}$, and $\vec{\mathbf{v}}$ but, when these are combined, all observers form the same assessment of $\vec{\mathbf{F}}/q$, and all obtain the same value for the induced emf in the loop (which depends only on the *relative* motion). That is, the total force (and, hence, the total acceleration) is the same for all observers, but each observer forms a different estimate of the separate electric and magnetic forces contributing to the same total force.

The essential point is that what seems like a magnetic field to one observer may seem like a mixture of an electric field and a magnetic field to a second observer in a different inertial reference frame. Both observers agree, however, on the overall measurable result—in the case of Fig. 34-19, the current in the loop. We are forced to conclude that magnetic and electric fields are *not* independent of each other and have no separate unique existence; they depend on the inertial frame, as we also concluded in Section 33-6.

All the results of this section assume that the relative speed between S and S' is small compared with the speed of light c. If v is comparable to c, the appropriate set of relativistic transformations must be applied. In this case, we would find that the induced emfs measured by S and S' would no longer be equal, and that the induced electric field is not given by Eq. 34-27. However, if we are careful to define all quantities in the proper relativistic manner, we find again that the basic laws of electromagnetism, including Faraday's law, hold in all inertial reference frames.* Indeed, such considerations led Einstein to the special theory of relativity; in the language of special relativity, we say that Maxwell's equations are invariant with respect to the Lorentz transformation. ■

---

*For a careful discussion of motional emfs in the case of velocities that are not necessarily small compared with c, see "Application of Special Relativity to a Simple System in which a Motional emf Exists," by Murray D. Sirkis, *American Journal of Physics*, June 1986, p. 538. Further considerations of the relativistic transformation of electric and magnetic fields can be found in *Introduction to Special Relativity*, by Robert Resnick (Wiley, 1968), Chapter 4.

MULTIPLE CHOICE

34-1 Faraday's Experiments

34-2 Faraday's Law of Induction

1. A uniform magnetic $\vec{\mathbf{B}}$ perpendicular to a wire loop varies with time as shown in Fig. 34-20*a*. Which of the plots in Fig. 34-20*b* best represents the induced current in the loop as a function of time?

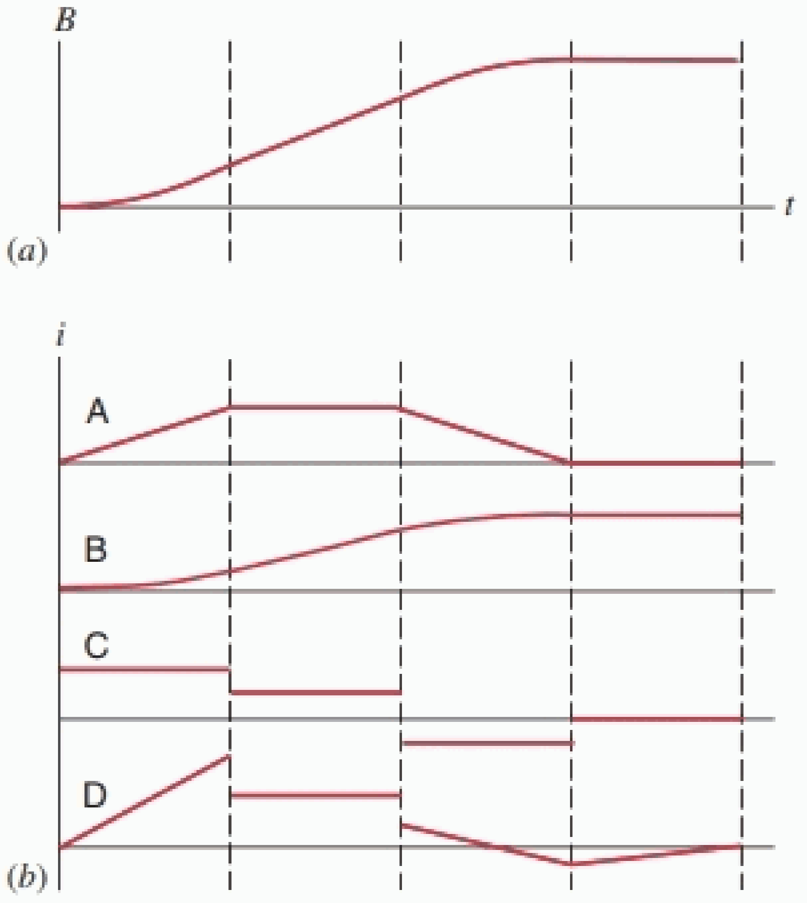


FIGURE 34-20. Multiple-choice question 1.

2. A flexible conducting loop is in the shape of a circle with a variable radius. The loop is in a uniform magnetic field perpendicular to the plane of the loop. To sustain a constant emf $\mathcal{E}$ in the loop, the radius r must vary with time according to

(A) $r(t) \propto \sqrt{t}$. (B) $r(t) \propto t$. (C) $r(t) \propto t^2$.
(D) r should be constant.

3. A flexible wire loop in the shape of a circle has a radius that grows linearly with time. There is a magnetic field perpendicular to the plane of the loop that has a magnitude inversely proportional to the distance from the center of loop, $B(r) \propto 1/r$. How does the emf $\mathcal{E}$ vary with time?

(A) $\mathcal{E} \propto t^2$ (B) $\mathcal{E} \propto t$ (C) $\mathcal{E} \propto \sqrt{t}$
(D) $\mathcal{E}$ is constant.

4. The magnetic flux through a wire loop changes by $\Delta\Phi_B$ in a time Δt. The change in flux $\Delta\Phi_B$ is proportional to

(A) the current in the wire.
(B) the resistance of the wire.
(C) the net charge that flows across any cross section in the wire.
(D) the potential difference between any two fixed points in the wire.

34-3 Lenz' Law

5. A bar magnet drops through a wire loop. Which graph in Fig. 34-21 best shows how the current through the loop varies with time? Assume that a negative current refers to a current flowing in the reverse direction.

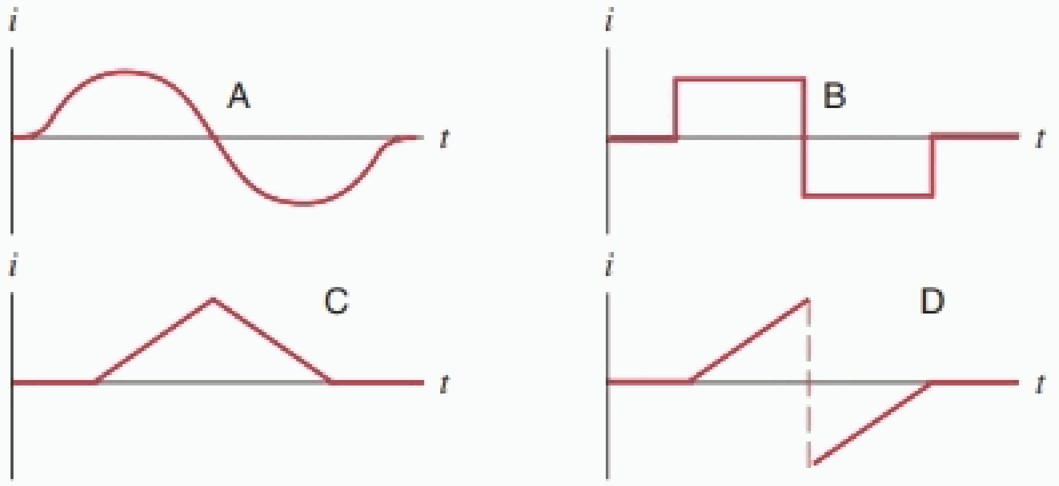


FIGURE 34-21. Multiple-choice question 5.

6. The current through the coil on the right in Fig. 34-2 varies as shown in Fig. 34-22*a*. Which graph in Fig. 34-22*b* best shows the ammeter reading as a function of time?

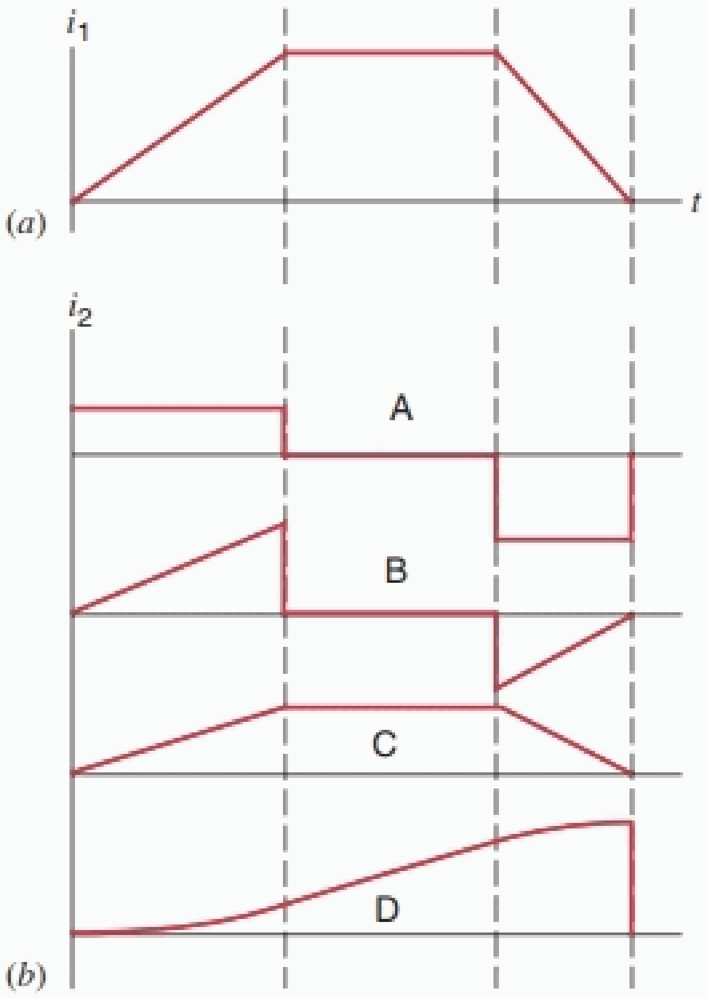


FIGURE 34-22. Multiple-choice question 6.

7. The long, straight wire in Fig. 34-23 carries a constant current i to the right. What is the direction of the induced current in the wire loop?

(A) Clockwise (B) Counterclockwise
(C) There is no induced current.

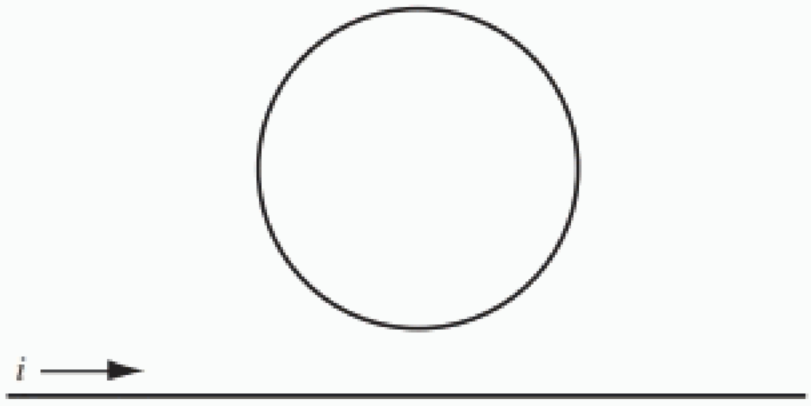


FIGURE 34-23. Multiple-choice questions 7, 8, 9, and 11.

8. The long, straight wire in Fig. 34-23 carries a current i to the right that increases with time. What is the direction of the induced current in the wire loop?

(A) Clockwise (B) Counterclockwise
(C) There is no induced current.

9. The long, straight wire in Fig. 34-23 carries a current i to the right that decreases linearly with time. At time t_r the current is zero, and then begins to increase linearly in the other direction. What is the direction of the induced current in the wire loop?

(A) Clockwise (B) Counterclockwise
(C) The induced current starts off in one direction, but at time t_r it stops and then begins flowing in the other direction.
(D) There is no induced current.

10. The long, straight wire in Fig. 34-24 carries a current i to the right that increases with time. What is the direction of the induced current in the circular wire loop, which lies in a plane perpendicular to the wire?

(A) Clockwise as viewed from the right
(B) Counterclockwise as viewed from the right
(C) There is no induced current.

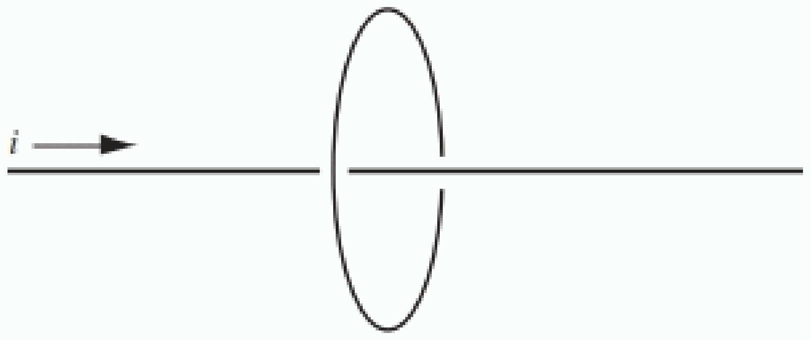


FIGURE 34-24. Multiple-choice question 10.

34-4 Motional EMF

11. Consider again the straight wire and loop of Fig. 34-23. The straight wire carries a constant current i to the right.

(*a*) The loop is pulled directly up by an external force. What is the direction (if any) of the induced current in the loop?

(A) Clockwise (B) Counterclockwise
(C) There is no induced current.

(*b*) While the loop is being pulled upward by the external force, what (if any) is the direction of the magnetic force on the loop?

(A) Up (B) Down (C) Left
(D) There is no magnetic force.

(*c*) If the external force instead pulls the loop to the left, parallel to the wire, what is the direction (if any) of the induced current in the loop?

(A) Clockwise (B) Counterclockwise
(C) There is no induced current.

34-5 Generators and Motors

12. A person uses a hand-cranked generator to light a constant-resistance bulb. As the person's rate of cranking slows down, which of the following quantities decreases?

(A) Emf (B) Current (C) Frequency
(D) Two of the above
(E) All of the above

13. Consider a bicycle generator attached to a wheel and connected to a constant-resistance light. If the speed of pedaling doubles, the power output from the light will

(A) remain the same. (B) double.
(C) increase by a factor of four.
(D) increase by a factor of eight.

14. Does an electric motor have a higher effective resistance when it is stopped or when it is spinning?

(A) When spinning (B) When stopped
(C) The effective resistance is the same in both cases.

15. All other things being equal, are you more likely to "burn out" the starter motor on your automobile if the battery is overcharged or undercharged?

(A) If the battery is overcharged
(B) If the battery is undercharged
(C) The charge of the battery makes no difference here.

34-6 Induced Electric Fields

16. The magnetic field in a region of space is given by $\vec{\mathbf{B}} = (0.001\ \mathrm{T/s^2})t^2\hat{\mathbf{i}}$ for $-2\ \mathrm{s} \le t \le 2\ \mathrm{s}$. What is the direction of the induced electric field when $t = 0$ s?

(A) Parallel to the x axis (B) Parallel to the y axis
(C) The electric field is in circles centered on the x axis.
(D) There is no induced electric field when $t = 0$ s.

17. The current through an infinitely long solenoid is increased linearly as a function of time.

(*a*) The electric field inside the solenoid is

(A) in the form of circles centered on the axis of the solenoid.
(B) parallel to the axis of the solenoid.
(C) radially directed outward from the axis of the cylinder.
(D) zero.

(*b*) The magnitude of the electric field outside the solenoid is

(A) uniform and nonzero.
(B) radially symmetric, falling off with distance from the solenoid.
(C) radially symmetric, increasing with the distance from the solenoid.
(D) zero.

18. A long, straight wire carries a current that decreases linearly with time. What is the direction of the induced electric field outside the wire?

(A) Parallel to the current
(B) Opposite to the current
(C) Pointing radially outward from the wire
(D) Pointing radially inward toward the wire
(E) There is no induced electric field outside the wire.

34-7 Induction and Relative Motion

QUESTIONS

1. Show that 1 volt = 1 weber/second.

2. Are induced emfs and currents different in any way from emfs and currents provided by a battery connected to a conducting loop?

3. Is the size of the voltage induced in a coil through which a magnet moves affected by the strength of the magnet? If so, explain how.

4. Explain in your own words the difference between a magnetic field $\vec{\mathbf{B}}$ and the flux of a magnetic field Φ_B. Are they vectors or scalars? In what units may each be expressed? How are their units related? Are either or both (or neither) properties of a given point in space?

5. Can a charged particle at rest be set in motion by the action of a magnetic field? If not, why not? If so, how? Consider both static and time-varying fields.

6. In Faraday's law of induction, does the induced emf depend on the resistance of the circuit? If so, how?

7. You drop a bar magnet along the axis of a long copper tube. Describe the motion of the magnet and the energy interchanges involved. Neglect air resistance.

8. You are playing with a metal loop, moving it back and forth in a magnetic field, as in Fig. 34-9. How can you tell, without detailed inspection, whether or not the loop has a narrow saw cut across it that is rendering it nonconducting?

9. Figure 34-25 shows an inclined wooden track that passes, for part of its length, through a strong magnetic field. You roll a copper disk down the track. Describe the motion of the disk as it rolls from the top of the track to the bottom.

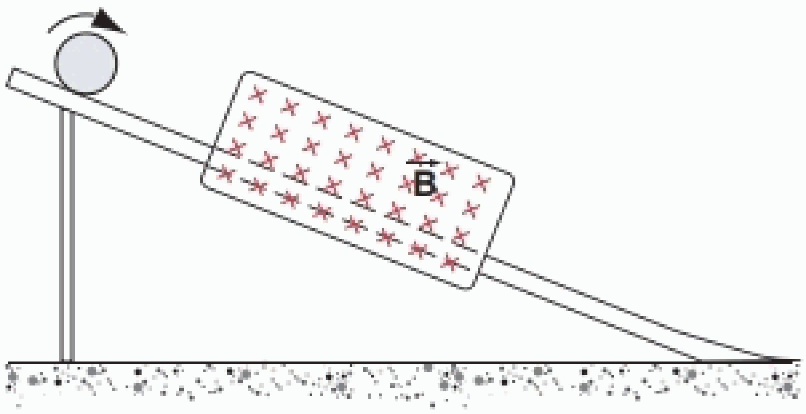


FIGURE 34-25. Question 9.

10. Figure 34-26 shows a copper ring that is hung from a ceiling by two threads. Describe in detail how you might most effectively use a bar magnet to get this ring to swing back and forth.

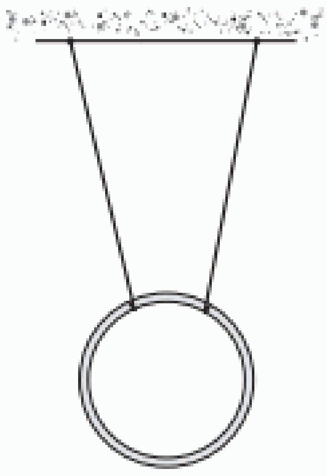


FIGURE 34-26. Question 10.

11. Is an emf induced in a long solenoid by a bar magnet that moves inside it along the solenoid axis? Explain your answer.

12. Two conducting loops face each other a distance d apart, as shown in Fig. 34-27. An observer sights along their common axis from left to right. A clockwise current i is suddenly established in the larger loop by a battery not shown. (*a*) What is the direction of the induced current in the smaller loop? (*b*) What is the direction of the force (if any) that acts on the smaller loop?

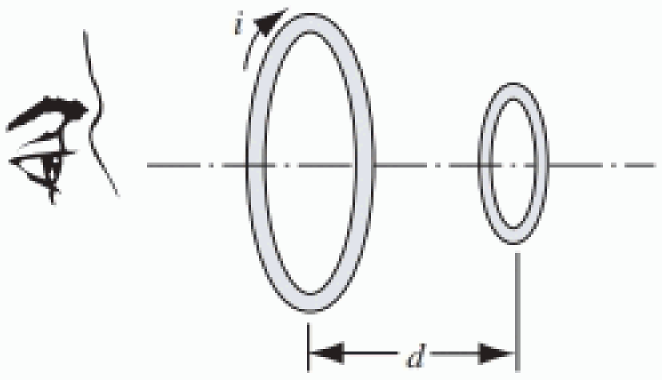


FIGURE 34-27. Question 12.

13. What is the direction of the induced current in coil Y of Fig. 34-28 (*a*) when coil Y is moved toward coil X? (*b*) When the current in coil X is decreased, without any change in the relative positions of the coils?

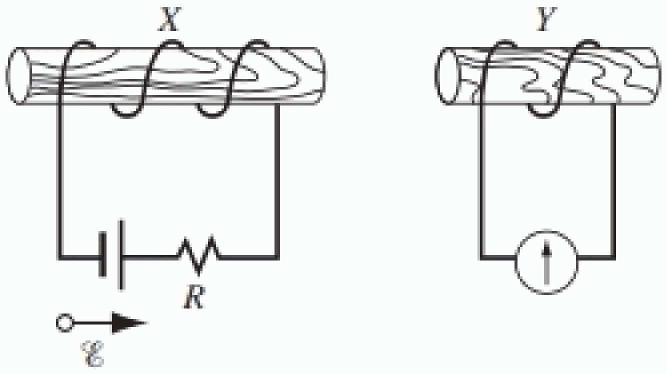


FIGURE 34-28. Question 13.

14. The north pole of a magnet is moved away from a copper ring, as in Fig. 34-29. In the part of the ring farthest from the reader, what is the direction of the current?

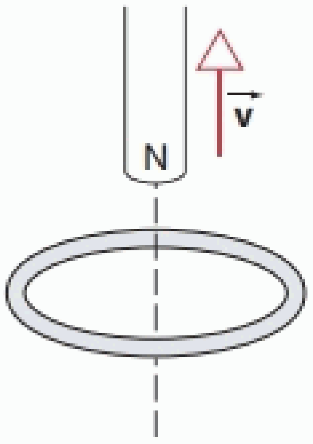


FIGURE 34-29. Question 14.

15. A circular loop moves with constant velocity through regions where uniform magnetic fields of the same magnitude are directed into or out of the plane of the page, as indicated in Fig. 34-30. At which of the seven indicated positions will the induced current be (*a*) clockwise, (*b*) counterclockwise, and (*c*) zero?

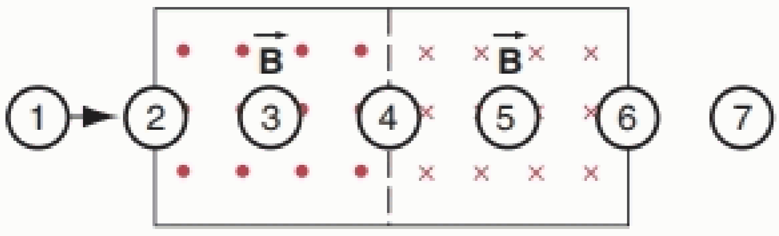


FIGURE 34-30. Question 15.

16. A short solenoid carrying a steady current is moving toward a conducting loop as in Fig. 34-31. What is the direction of the induced current in the loop as one sights toward it as shown?

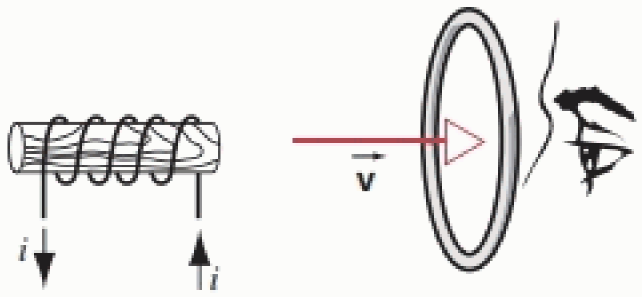


FIGURE 34-31. Question 16.

17. The resistance *R* in the left-hand circuit of Fig. 34-32 is being increased at a steady rate. What is the direction of the induced current in the right-hand circuit?

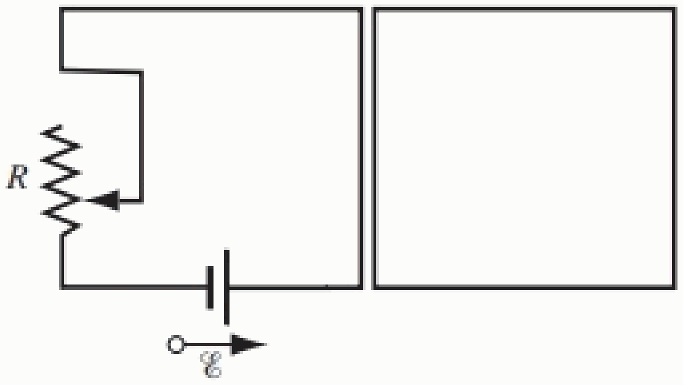


FIGURE 34-32. Question 17.

18. What is the direction of the induced current through resistor *R* in Fig. 34-33 (*a*) immediately after switch S is closed, (*b*) some time after switch S is closed, and (*c*) immediately after switch S is opened? (*d*) When switch S is held closed, from which end of the longer coil do field lines emerge? This is the effective north pole of the coil. (*e*) How do the conduction electrons in the coil containing *R* know about the flux within the long coil? What really gets them moving?

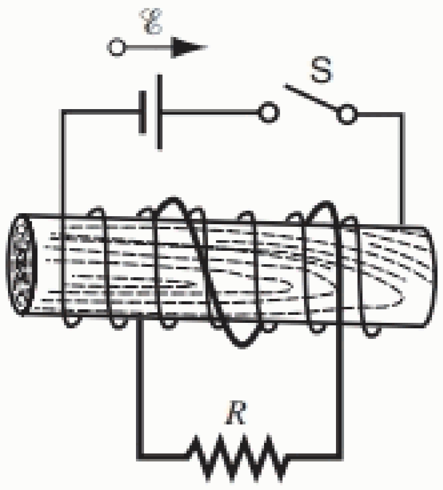


FIGURE 34-33. Question 18.

19. Can an induced current ever establish a magnetic field $\vec{B}$ that is in the same direction as the magnetic field inducing the current? Justify your answer.

20. How can you summarize in one statement all the ways of determining the direction of an induced emf?

21. The loop of wire shown in Fig. 34-34 rotates with constant angular speed about the *x* axis. A uniform magnetic field $\vec{B}$, whose direction is that of the positive *y* axis, is present. For what portions of the rotation is the induced current in the loop (*a*) from *P* to *Q*, (*b*) from *Q* to *P*, and (*c*) zero? Repeat if the direction of rotation is reversed from that shown in the figure.

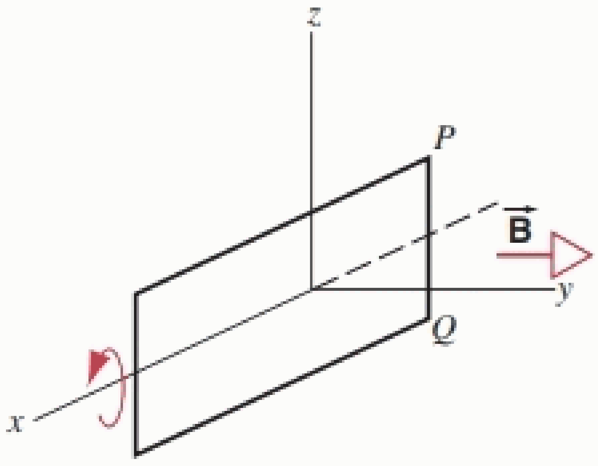


FIGURE 34-34. Question 21.

22. In Fig. 34-35, the straight movable wire segment is moving to the right with constant velocity $\vec{v}$. An induced current appears in the direction shown. What is the direction of the uniform magnetic field (assumed to be constant and perpendicular to the page) in region *A*?

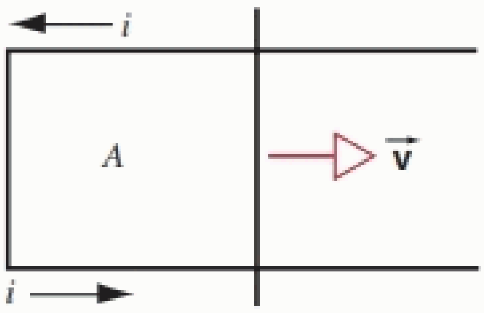


FIGURE 34-35. Question 22.

23. A conducting loop, shown in Fig. 34-36, is removed from the permanent magnet by pulling it vertically upward. (*a*) What is the direction of the induced current? (*b*) Is a force required to remove the loop? (Ignore the weight of the loop.) (*c*) Does the total amount of internal energy produced depend on the time taken to remove it?

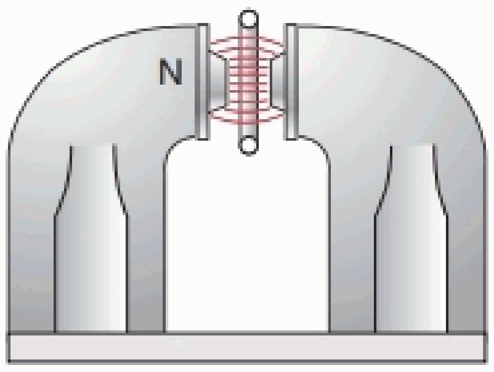


FIGURE 34-36. Question 23.

24. A plane closed loop is placed in a uniform magnetic field. In what ways can the loop be moved without inducing an emf? Consider motions both of translation and rotation.

25. A strip of copper is mounted as a pendulum about O in Fig. 34-37. It is free to swing through a magnetic field that is normal to the page. If the strip has slots cut in it as shown, it can swing freely through the field. If a strip without slots is substituted, the motion is strongly damped (*magnetic damping*). Explain the observations. (Hint: Use Lenz' law; consider the paths that the charge carriers in the strip must follow if they are to oppose the motion.)

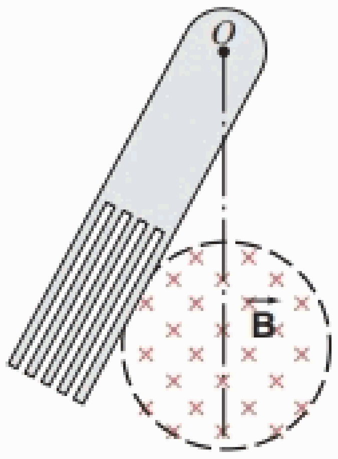


FIGURE 34-37. Question 25.

26. Consider a conducting sheet lying in a plane perpendicular to a magnetic field $\vec{B}$, as shown in Fig. 34-38. (*a*) If $\vec{B}$ suddenly changes, the full change in $\vec{B}$ is not immediately detected at points near P (electromagnetic shielding). Explain. (*b*) If the resistivity of the sheet is zero, the change is never detected at P. Explain. (*c*) If $\vec{B}$ changes periodically at high frequency and the conductor is made of material with a low resistivity, the region near P is almost completely shielded from the changes in flux. Explain. (*d*) Why is such a conductor not useful as a shield from static magnetic fields?

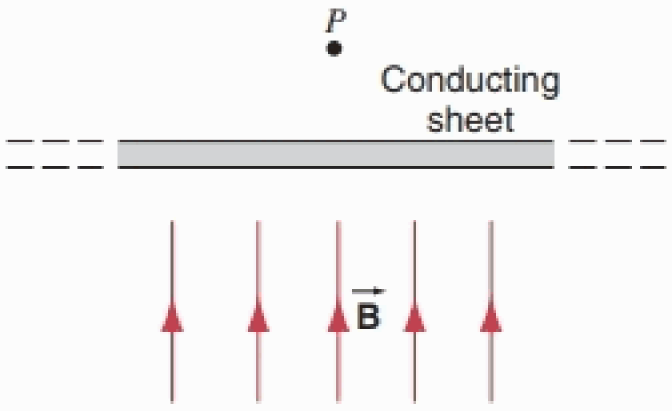


FIGURE 34-38. Question 26.

27. (*a*) In Fig. 34-15*b*, need the circle of radius r be a conducting loop in order that $\vec{E}$ and $\mathcal{E}$ be present? (*b*) If the circle of radius r were not concentric (moved slightly to the left, say), would $\mathcal{E}$ change? Would the configuration of $\vec{E}$ around the circle change? (*c*) For a concentric circle of radius r, with $r > R$, does an emf exist? Do electric fields exist?

28. A copper ring and a wooden ring of the same dimensions are placed so that there is the same changing magnetic flux through each. Compare the induced electric fields in the two rings.

29. An airliner is cruising in level flight over Alaska, where Earth's magnetic field has a large downward component. Which of its wingtips (right or left) has more electrons than the other?

30. In Fig. 34-15*d*, how can the induced emfs around paths 1 and 2 be identical? The induced electric fields are much weaker near path 1 than near path 2, as the spacing of the field lines shows. See also Fig. 34-17.

31. Show that, in the betatron of Fig. 34-18, the directions of the lines of $\vec{B}$ are correctly drawn to be consistent with the direction of circulation shown for the electrons.

32. Figure 34-39*a* shows a top view of the electron orbit in a betatron. Electrons are accelerated in a circular orbit in the xy plane and then withdrawn to strike the target T. The magnetic field $\vec{B}$ is along the z axis (the positive z axis is out of the page). The magnetic field B_z along this axis varies sinusoidally as shown in Fig. 34-39*b*. Recall that the magnetic field must (i) guide the electrons in their circular path and (ii) generate the electric field that accelerates the electrons. Which quarter cycle(s) in Fig. 34-39*b* are suitable (*a*) according to (i), (*b*) according to (ii), and (*c*) for operation of the betatron?

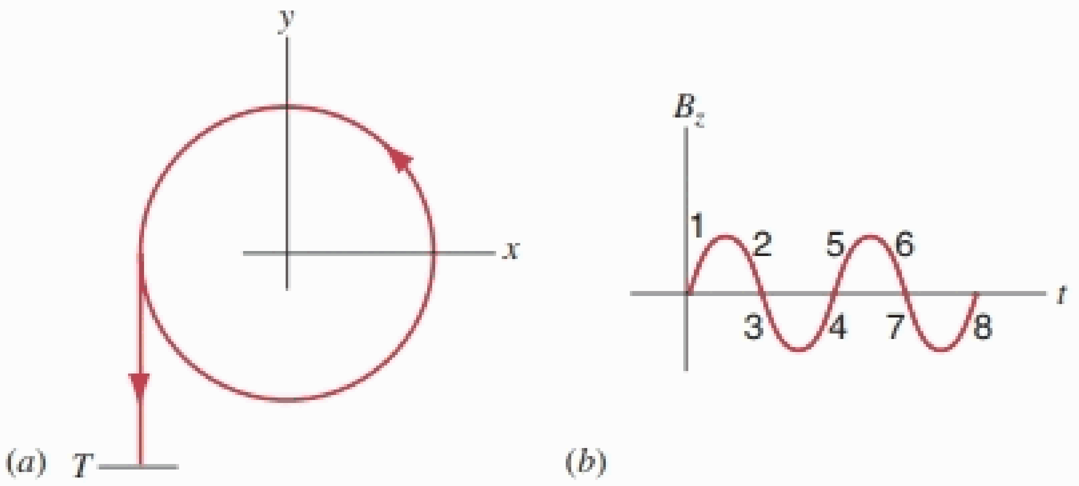


FIGURE 34-39. Question 32.

33. In the betatron of Fig. 34-18, why is the iron core of the magnet made of laminated sheets rather than of solid metal as for the cyclotron of Section 32-3?

34. In Fig. 34-19*a* we can see that a force ($F_B \cos \theta$) acts on the charge carriers in the left branch of the loop. However, if there is to be a continuous current in the loop, and there is, a force of some sort must act on charge carriers in the other three branches of the loop to maintain the same drift speed v_d in these branches. What is its source? (Hint: Consider that the left branch of the loop was the only conducting element, the other three being nonconducting. Would not positive charge pile up at the top of the left half and negative charge at the bottom?)

EXERCISES

34-1 Faraday's Experiments

34-2 Faraday's Law of Induction

1. At a certain location in the northern hemisphere, the Earth's magnetic field has a magnitude of 42 μT and points downward at 57° to the vertical. Calculate the flux through a horizontal surface of area 2.5 m^2; see Fig. 34-40.

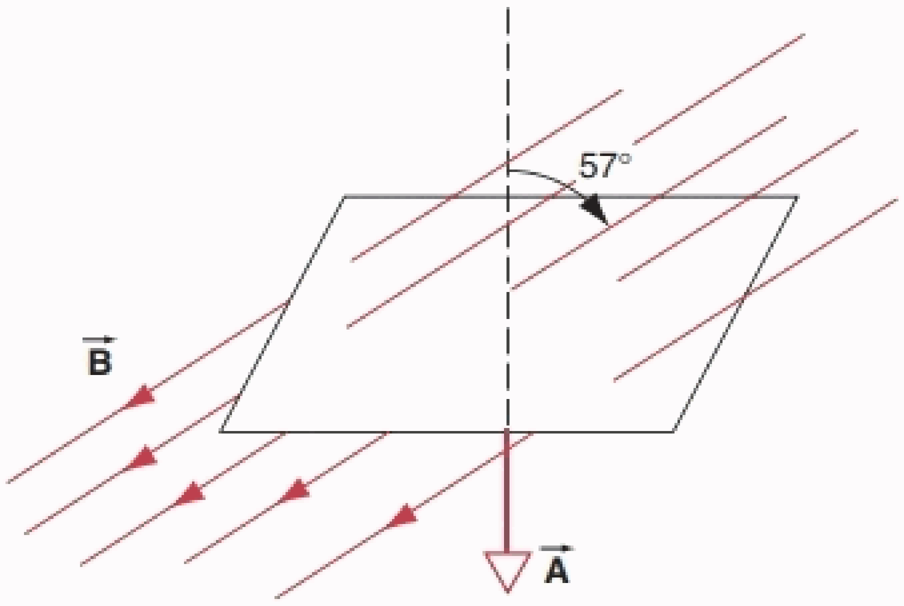


FIGURE 34-40. Exercise 1.

2. A circular UHF television antenna has a diameter of 11.2 cm. The magnetic field of a TV signal is normal to the plane of the loop and, at one instant of time, its magnitude is changing at the rate 157 mT/s. The field is uniform. Find the emf in the antenna.

34-3 Lenz' Law

3. In Fig. 34-41, the magnetic flux through the loop shown increases according to the relation

$$\Phi_B = (6\ \text{mWb/s}^2)t^2 + (7\ \text{mWb/s})t.$$

(*a*) What is the absolute value of the emf induced in the loop when $t = 2.0$ s? (*b*) What is the direction of the current through the resistor?

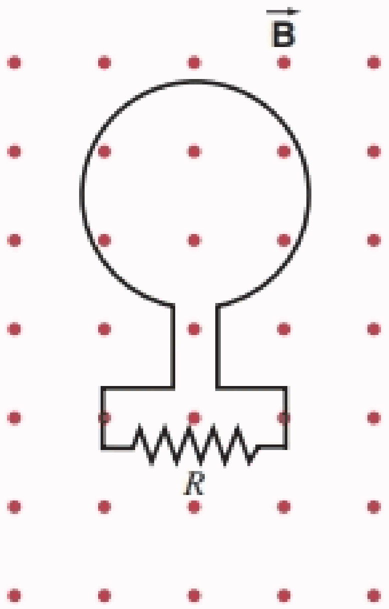


FIGURE 34-41. Exercises 3 and 11.

4. The magnetic field through a one-turn loop of wire 16 cm in radius and 8.5 Ω in resistance changes with time as shown in Fig. 34-42. Calculate the emf in the loop as a function of time. Consider the time intervals (*a*) $t = 0$ to $t = 2$ s; (*b*) $t = 2$ s to $t = 4$ s; (*c*) $t = 4$ s to $t = 8$ s. The (uniform) magnetic field is perpendicular to the plane of the loop.

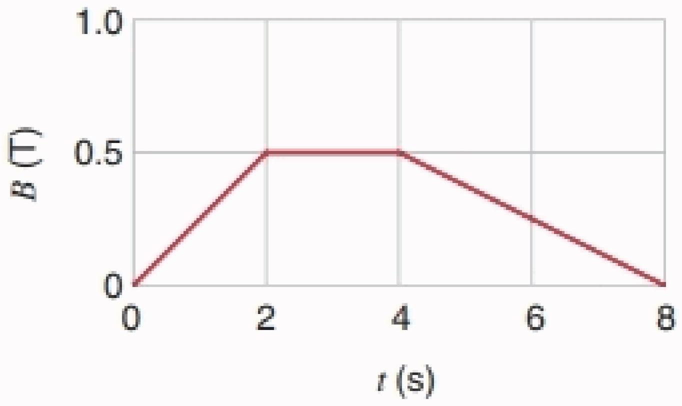


FIGURE 34-42. Exercise 4.

5. A uniform magnetic field is perpendicular to the plane of a circular loop 10.4 cm in diameter made of copper wire (diameter = 2.50 mm). (*a*) Calculate the resistance of the wire. (See Table 29-1.) (*b*) At what rate must the magnetic field change with time if an induced current of 9.66 A is to appear in the loop?
6. A loop antenna of area *A* and resistance *R* is perpendicular to a uniform magnetic field $\vec{B}$. The field drops linearly to zero in a time interval Δt. Find an expression for the total internal energy dissipated in the loop.
7. Suppose that the current in the solenoid of Sample Problem 34-1 now changes, not as in that sample problem, but according to $i = (3.0\ \text{A/s})t + (1.0\ \text{A/s}^2)t^2$. (*a*) Plot the induced emf in the coil from $t = 0$ to $t = 4$ s. (*b*) The resistance of the coil is 0.15 Ω. What is the current in the coil at $t = 2.0$ s?
8. In Fig. 34-43, a 120-turn coil of radius 1.8 cm and resistance 5.3 Ω is placed outside a solenoid like that of Sample Problem 34-1. If the current in the solenoid is changed as in that sample problem, (*a*) what current appears in the coil while the solenoid current is being changed? (*b*) How do the conduction electrons in the coil "get the message" from the solenoid that they should move to establish a current? After all, the magnetic flux is entirely confined to the interior of the solenoid.

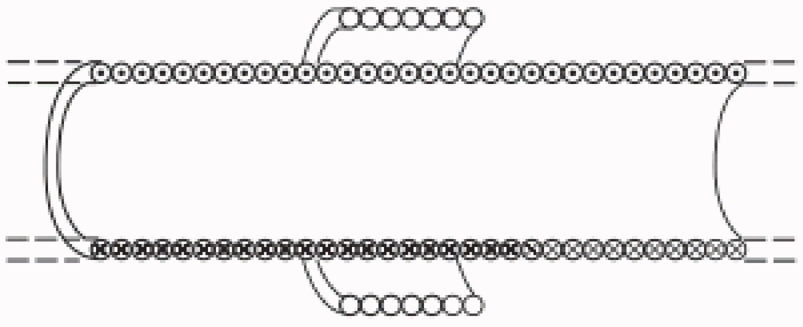


FIGURE 34-43. Exercise 8.

9. You are given 52.5 cm of copper wire (diameter = 1.10 mm). It is formed into a circular loop and placed at right angles to a uniform magnetic field that is increasing with time at the constant rate of 9.82 mT/s. At what rate is internal energy generated in the loop?
10. A square wire loop with 2.3-m sides is perpendicular to a uniform magnetic field, with half the area of the loop in the field, as shown in Fig. 34-44. The loop contains a 2.0-V battery

with negligible internal resistance. If the magnitude of the field varies with time according to $B = (0.042\text{ T}) - (0.87\text{ T/s})t$, what is the total emf in the circuit?

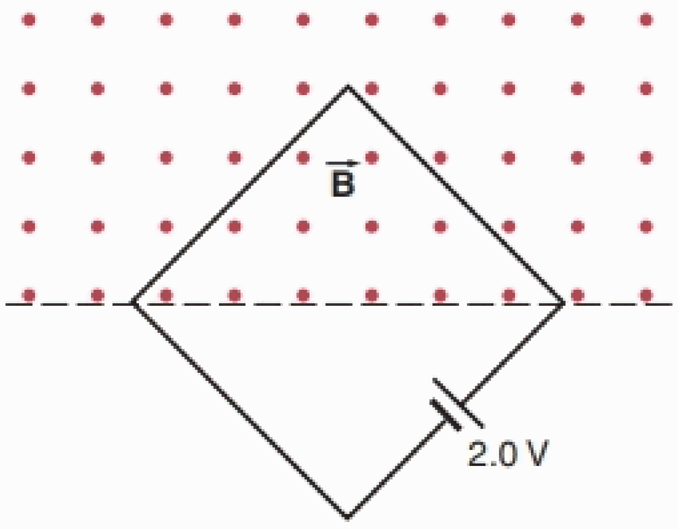


FIGURE 34-44. Exercise 10.

11. In Fig. 34-41, let the flux for the loop be $\Phi_B(0)$ at time $t = 0$. Then let the magnetic field $\vec{B}$ vary in a continuous but unspecified way, in both magnitude and direction, so that at time t the flux is represented by $\Phi_B(t)$. (*a*) Show that the net charge $q(t)$ that has passed through resistor R in time t is

$$q(t) = \frac{1}{R}[\Phi_B(0) - \Phi_B(t)],$$

independent of the way $\vec{B}$ has changed. (*b*) If $\Phi_B(t) = \Phi_B(0)$ in a particular case we have $q(t) = 0$. Is the induced current necessarily zero throughout the time interval from 0 to t?

12. Around a cylindrical core of cross-sectional area 12.2 cm^2 are wrapped 125 turns of insulated copper wire. The two terminals are connected to a resistor. The total resistance in the circuit is 13.3 Ω. An externally applied, uniform, longitudinal magnetic field in the core changes from 1.57 T in one direction to 1.57 T in the opposite direction in 2.88 ms. How much charge flows through the circuit? (Hint: See Exercise 11.)

13. For the situation shown in Fig. 34-45, $a = 12$ cm, $b = 16$ cm. The current in the long, straight wire is given by $i = (4.5\text{ A/s}^2)t^2 - (10\text{ A/s})t$. Find the emf in the square loop at $t = 3.0$ s.

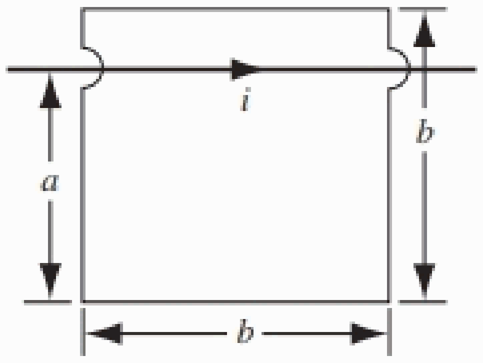


FIGURE 34-45. Exercise 13.

34-4 Motional EMF

14. An automobile having a radio antenna 110 cm long travels at 90 km/h in a region where Earth's magnetic field is 55 μT. Find the maximum possible value of the induced emf.

15. A circular loop of wire 10 cm in diameter is placed with its normal making an angle of 30° with the direction of a uniform 0.50 T magnetic field. The loop is "wobbled" so that its normal rotates in a cone about the field direction at the constant rate of 100 rev/min; the angle between the normal and the field direction (= 30°) remains unchanged during the process. What emf appears in the loop?

16. Figure 34-46 shows a conducting rod of length L being pulled along horizontal, frictionless, conducting rails at a constant velocity $\vec{v}$. A uniform vertical magnetic field $\vec{B}$ fills the region in which the rod moves. Assume that $L = 10.8$ cm, $v = 4.86$ m/s, and $B = 1.18$ T. (*a*) Find the induced emf in the rod. (*b*) Calculate the current in the conducting loop. Assume that the resistance of the rod is 415 mΩ and that the resistance of the rails is negligibly small. (*c*) At what rate does the internal energy of the rod increase? (*d*) Find the force that must be applied by an external agent to the rod to maintain its motion. (*e*) At what rate does this force do work on the rod? Compare this answer with the answer to (*c*).

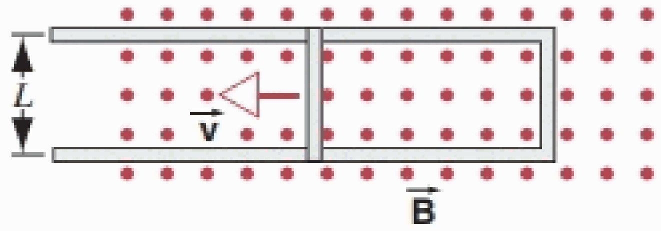


FIGURE 34-46. Exercise 16.

17. In Fig. 34-47, a conducting rod of mass m and length L slides without friction on two long, horizontal rails. A uniform vertical magnetic field $\vec{B}$ fills the region in which the rod is free to move. The generator G supplies a current i that flows down one rail, across the rod, and back to the generator along the other rail. A student monitors the generator, continually adjusting it so that the current it supplies is constant regardless of the load. Find the velocity of the rod as a function of time, assuming it to be at rest at $t = 0$.

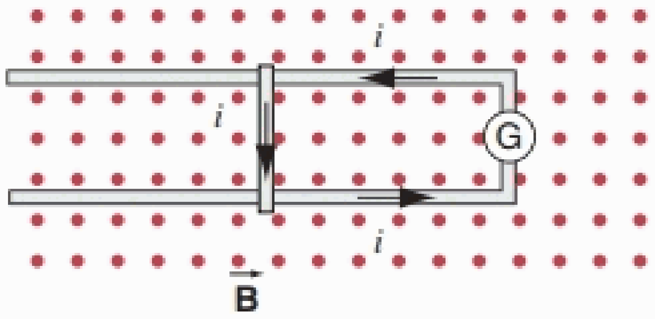


FIGURE 34-47. Exercises 17 and 18.

18. In Exercise 17 (see Fig. 34-47) the generator G is replaced by a battery that supplies a constant emf $\mathcal{E}$. (*a*) Show that the velocity of the rod now approaches a constant terminal value $\vec{v}$ and give its magnitude and direction. (*b*) What is the current in the rod when this terminal velocity is reached? (*c*) Analyze both this situation and that of Exercise 17 from the point of view of energy transfers.

19. A small bar magnet is pulled rapidly through a conducting loop, along its axis. Sketch qualitatively (*a*) the induced current and (*b*) the rate of internal energy production as a function of the position of the center of the magnet. Assume that the north pole of the magnet enters the loop first and that the magnet moves at constant speed. Plot the induced current as positive if it is clockwise as viewed along the path of the magnet.

20. In the arrangement of Sample Problem 34-4, put $B = 1.2$ T and $R = 5.3$ cm. If $\mathcal{E} = 1.4$ V, what acceleration will a point at the end of the rotating rod experience?

21. At a certain place, the Earth's magnetic field has magnitude $B = 59\ \mu\text{T}$ and is inclined downward at an angle of 70° to the horizontal. A flat, horizontal, circular coil of wire with a radius of 13 cm has 950 turns and a total resistance of 85 Ω. The coil is flipped through a half revolution about a diameter, so it is again horizontal. How much charge flows through the coil during the flip? (Hint: See Exercise 11.)

22. Figure 34-48 shows a rod of length L caused to move at constant speed v along horizontal conducting rails. In this case the magnetic field in which the rod moves is not uniform but is provided by a current i in a long, parallel wire. Assume that $v = 4.86$ m/s, $a = 10.2$ mm, $L = 9.83$ cm, and $i = 110$ A. (*a*) Calculate the emf induced in the rod. (*b*) What is the current in the conducting loop? Assume that the resistance of the rod is 415 mΩ and that the resistance of the rails is negligible. (*c*) At what rate does the internal energy of the rod increase? (*d*) What force must be applied to the rod by an external agent to maintain its motion? (*e*) At what rate does this external agent do work on the rod? Compare this answer to (*c*).

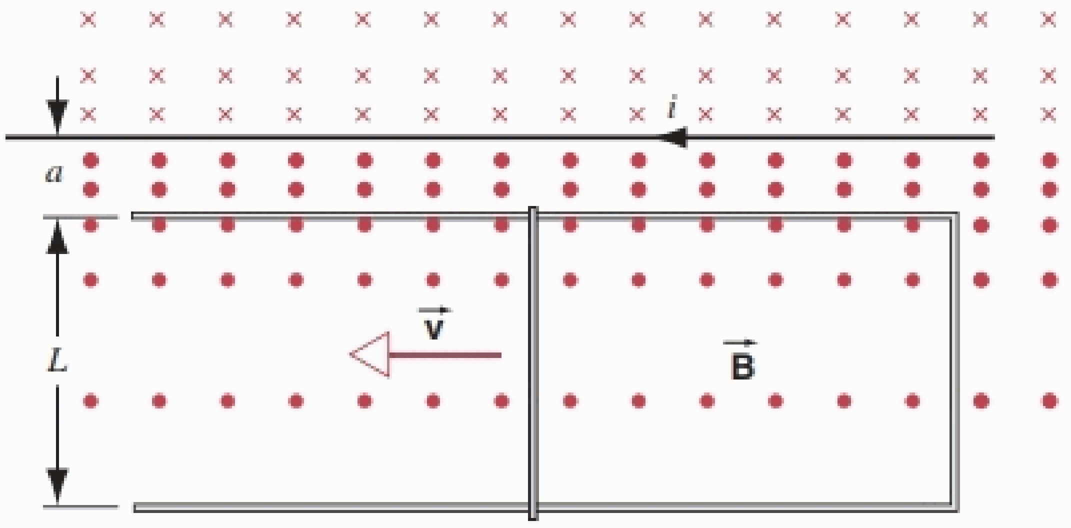


FIGURE 34-48. Exercise 22.

23. A rectangular loop of wire with length a, width b, and resistance R is placed near an infinitely long wire carrying current i, as shown in Fig. 34-49. The distance from the long wire to the loop is D. Find (*a*) the magnitude of the magnetic flux through the loop and (*b*) the current in the loop as it moves away from the long wire with speed v.

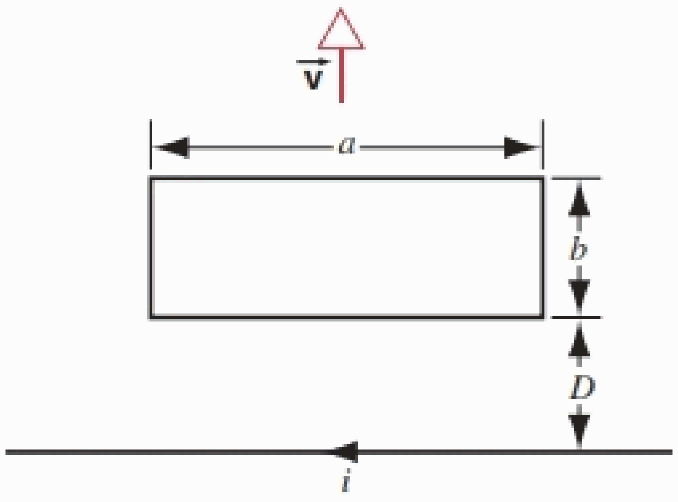


FIGURE 34-49. Exercise 23.

24. Two straight, conducting rails form an angle θ where their ends are joined. A conducting bar in contact with the rails and forming an isosceles triangle with them starts at the vertex at time $t = 0$ and moves with constant velocity $\vec{v}$ to the right, as shown in Fig. 34-50. A magnetic field $\vec{B}$ points out of the page. (*a*) Find the emf induced as a function of time. (*b*) If $\theta = 110°$, $B = 352$ mT, and $v = 5.21$ m/s, when is the induced emf equal to 56.8 V?

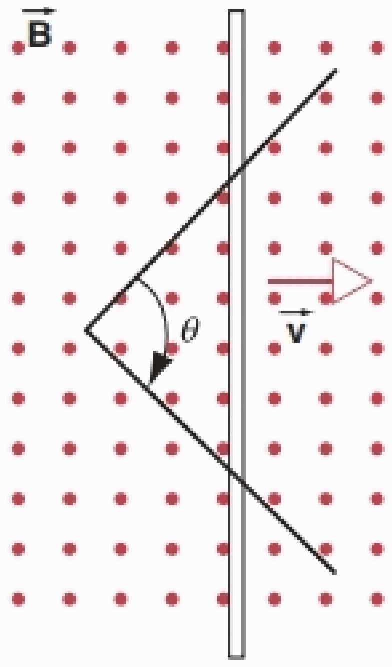


FIGURE 34-50. Exercise 24.

34-5 Generators and Motors

25. The armature of a motor has 97 turns, each of area 190 cm^2, and rotates in a uniform magnetic field of 0.33 T. A potential difference of 24 V is applied. If no load is attached and friction is neglected, find the rotational speed at equilibrium.

26. A stiff wire bent into a semicircle of radius a is rotated with a frequency f in a uniform magnetic field, as suggested in Fig. 34-51. What are (*a*) the frequency and (*b*) the amplitude of the emf induced in the loop?

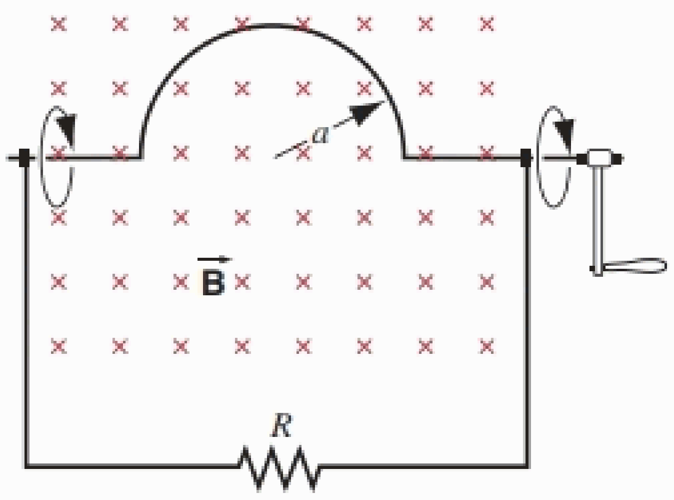


FIGURE 34-51. Exercise 26.

27. A generator consists of 100 turns of wire formed into a rectangular loop 50 cm by 30 cm, placed entirely in a uniform magnetic field with magnitude $B = 3.5$ T. What is the maximum value of the emf produced when the loop is spun at 1000 revolutions per minute about an axis perpendicular to $\vec{B}$?

28. You are assigned the job of designing a generator that will produce an emf of amplitude 150 V when rotated at 60 rev/s in a 0.50-T magnetic field. (*a*) If you used a single-turn loop, how large an area would you need? (*b*) If instead you used a loop of 100 turns, what area would be required?

34-6 Induced Electric Fields

29. Figure 34-52 shows two circular regions R_1 and R_2 with radii $r_1 = 21.2$ cm and $r_2 = 32.3$ cm, respectively. In R_1 there is a uniform magnetic field $B_1 = 48.6$ mT into the page and in R_2 there is a uniform magnetic field $B_2 = 77.2$ mT out of the page (ignore any fringing of these fields). Both fields are decreasing at the rate 8.50 mT/s. Calculate the integral $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}$ for each of the three indicated paths.

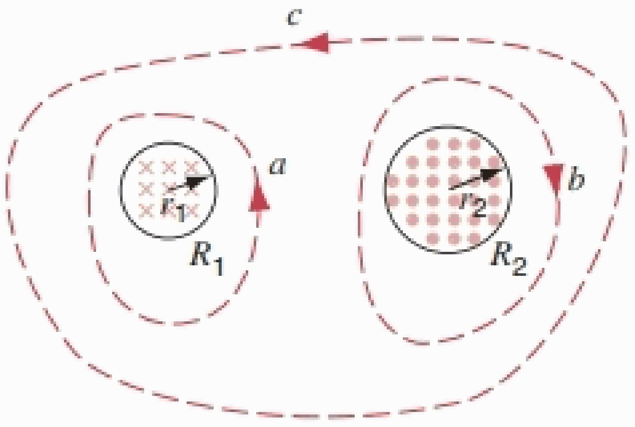


FIGURE 34-52. Exercise 29.

30. A long solenoid has a diameter of 12.6 cm. When a current i is passed through its windings, a uniform magnetic field $B =$ 28.6 mT is produced in its interior. By decreasing i, the field is caused to decrease at the rate 6.51 mT/s. Calculate the magnitude of the induced electric field (*a*) 2.20 cm and (*b*) 8.20 cm from the axis of the solenoid.

31. Figure 34-53 shows a uniform magnetic field $\vec{\mathbf{B}}$ confined to a cylindrical volume of radius R. $\vec{\mathbf{B}}$ is decreasing in magnitude at a constant rate of 10.7 mT/s. What is the instantaneous acceleration (direction and magnitude) experienced by an electron placed at a, at b, and at c? Assume that $r = 4.82$ cm. (The necessary fringing of the field beyond R will not change your answer as long as there is axial symmetry about the perpendicular axis through b.)

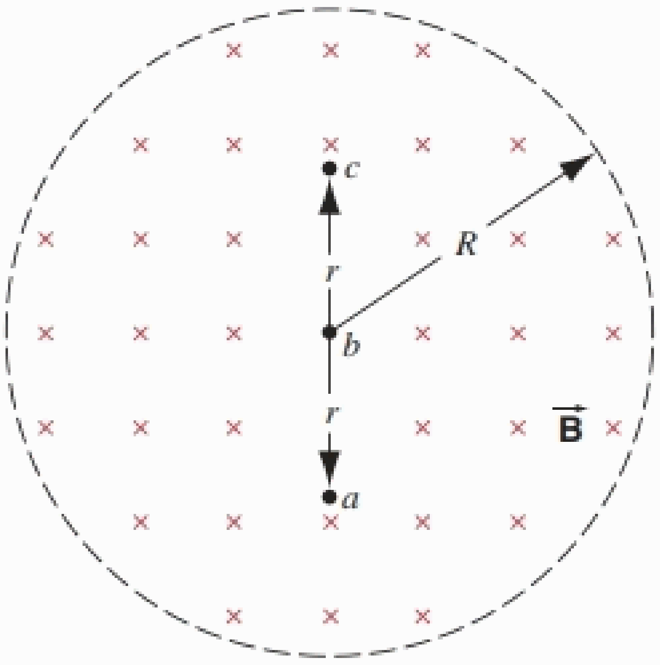


FIGURE 34-53. Exercise 31.

32. Early in 1981, the Francis Bitter National Magnet Laboratory at M.I.T. commenced operation of a 3.3-cm-diameter cylindrical magnet that produces a 30-T field, then the world's largest steady-state field. The field can be varied sinusoidally between the limits of 29.6 T and 30.0 T at a frequency of 15 Hz. When this is done, what is the maximum value of the induced electric field at a radial distance of 1.6 cm from the axis? This magnet is described in *Physics Today,* August 1984.

34-7 Induction and Relative Motion

33. (*a*) Estimate θ in Fig. 34-19. Recall that $v_d = 4 \times 10^{-2}$ cm/s in a typical case. Assume $v = 15$ cm/s. (*b*) It is clear that θ will be small. However, must we have $\theta \neq 0$ for the arguments presented in connection with this figure to be valid?

PROBLEMS

1. A uniform magnetic field $\vec{\mathbf{B}}$ is changing in magnitude at a constant rate dB/dt. You are given a mass m of copper that is to be drawn into a wire of radius r and formed into a circular loop of radius R. Show that the induced current in the loop does not depend on the size of the wire or of the loop and, assuming $\vec{\mathbf{B}}$ perpendicular to the loop, is given by

$$i = \frac{m}{4\pi\rho\delta}\frac{dB}{dt},$$

where ρ is the resistivity and δ the density of copper.

2. A closed loop of wire consists of a pair of identical semicircles, radius 3.7 cm, lying in mutually perpendicular planes. The loop was formed by folding a circular loop along a diameter until the two halves became perpendicular. A uniform magnetic field $\vec{\mathbf{B}}$ of magnitude 76 mT is directed perpendicular to the fold diameter and makes angles of 62° and 28° with the planes of the semicircles, as shown in Fig. 34-54. The magnetic field is reduced at a uniform rate to zero during a time interval of 4.5 ms. Determine the induced emf.

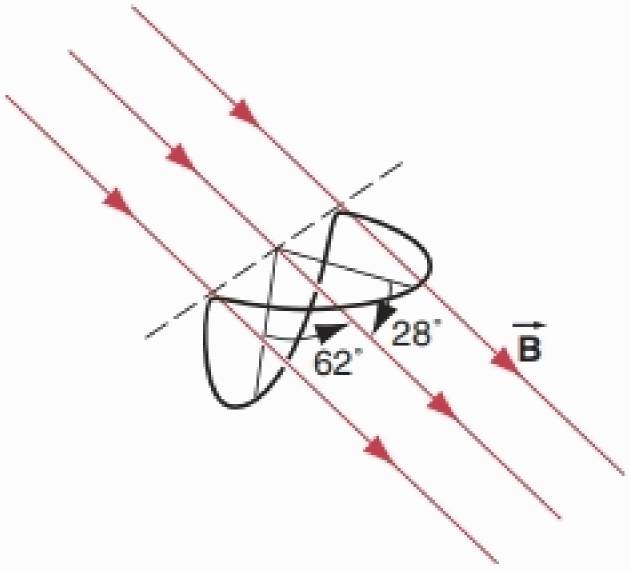


FIGURE 34-54. Problem 2.

3. A wire is bent into three circular segments of radius $r =$ 10.4 cm as shown in Fig. 34-55. Each segment is a quadrant of a circle, ab lying in the xy plane, bc lying in the yz plane, and ca lying in the zx plane. (*a*) If a uniform magnetic field $\vec{\mathbf{B}}$ points in the positive x direction, find the emf developed in the wire when B increases at the rate of 3.32 mT/s. (*b*) What is the direction of the current in the segment bc?

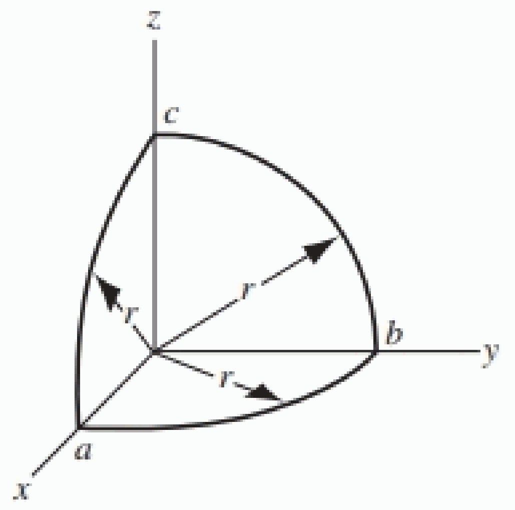


FIGURE 34-55. Problem 3.

4. A conducting wire of fixed length L can be wound into N circular turns and used as the armature of a generator. To get the largest emf, what value of N would you choose?

5. In Fig. 34-56, the square has sides of length 2.0 cm. A magnetic field points out of the page; its magnitude is given by $B = (4\ \mathrm{T/m \cdot s^2})t^2y$. Determine the emf around the square at $t = 2.5$ s and give its direction.

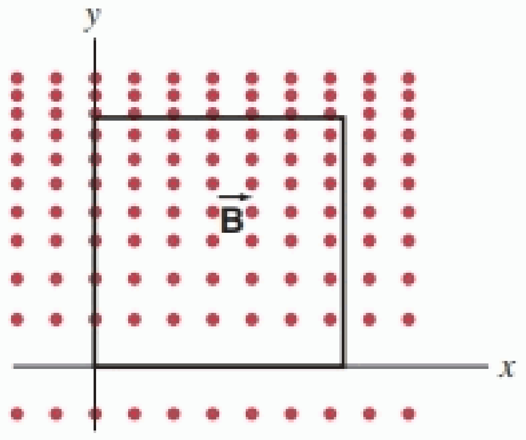


FIGURE 34-56. Problem 5.

6. Figure 34-57 shows two parallel loops of wire having a common axis. The smaller loop (radius r) is above the larger loop (radius R), by a distance $x \gg R$. Consequently the magnetic field, due to the current i in the larger loop, is nearly constant throughout the smaller loop and equal to the value on the axis. Suppose that x is increasing at the constant rate $dx/dt = v$. (*a*) Determine the magnetic flux across the area bounded by the smaller loop as a function of x. (*b*) Compute the emf generated in the smaller loop. (*c*) Determine the direction of the induced current flowing in the smaller loop.

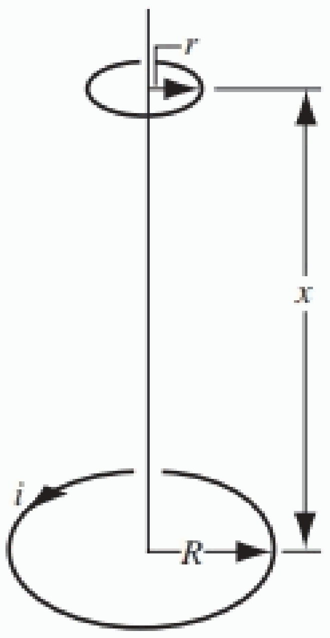


FIGURE 34-57. Problem 6.

7. A circular loop made of a stretched conducting elastic material has a 1.23-m radius. It is placed with its plane at right angles to a uniform 785-mT magnetic field. When released, the radius of the loop starts to decrease at an instantaneous rate of 7.50 cm/s. Calculate the emf induced in the loop at that instant.

8. Figure 34-58 shows a "homopolar generator," a device with a solid conducting disk as rotor. This machine can produce a greater emf than one using wire loop rotors, since it can spin at a much higher angular speed before centrifugal forces disrupt the rotor. (*a*) Show that the emf produced is given by

$$\mathcal{E} = \pi f B R^2,$$

where f is the spin frequency, R the rotor radius, and B the uniform magnetic field perpendicular to the rotor. (*b*) Find the torque that must be provided by the motor spinning the rotor when the output current is i.

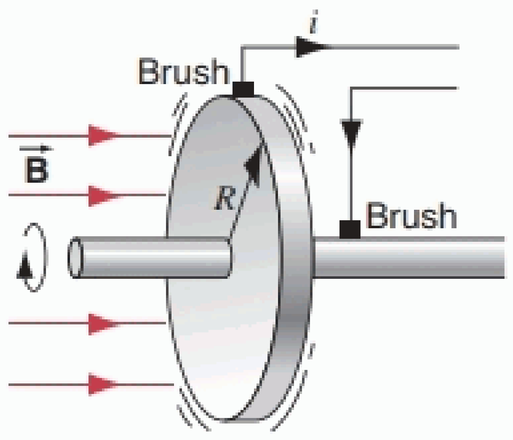


FIGURE 34-58. Problem 8.

9. A rod with length L, mass m, and resistance R slides without friction down parallel conducting rails of negligible resistance, as in Fig. 34-59. The rails are connected together at the bottom as shown, forming a conducting loop with the rod as the top member. The plane of the rails makes an angle θ with the horizontal, and a uniform vertical magnetic field $\vec{\mathbf{B}}$ exists throughout the region. (*a*) Show that the rod acquires a steady-state terminal velocity whose magnitude is

$$v = \frac{mgR}{B^2L^2}\frac{\sin\theta}{\cos^2\theta}.$$

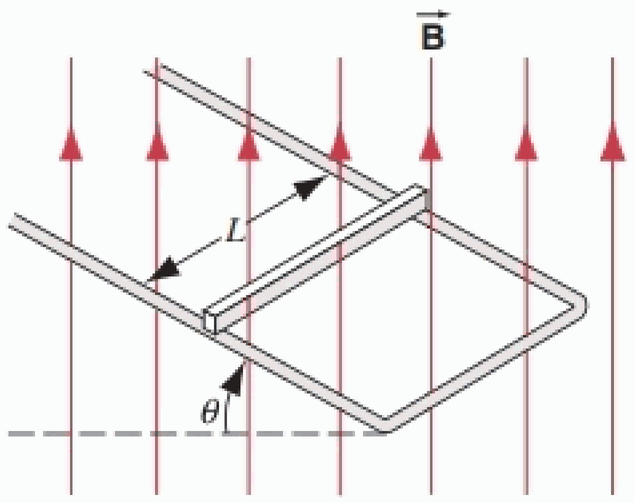


FIGURE 34-59. Problem 9.

(*b*) Show that the rate at which the internal energy of the rod is increasing is equal to the rate at which the rod is losing gravitational potential energy. (*c*) Discuss the situation if $\vec{\mathbf{B}}$ were directed down instead of up.

10. A wire whose cross-sectional area is 1.2 mm^2 and whose resistivity is 1.7×10^{-8} m is bent into a circular arc of radius $r = 24$ cm as shown in Fig. 34-60. An additional straight length of this wire, OP, is free to pivot about O and makes sliding contact with the arc at P. Finally, another straight length of this wire, OQ, completes the circuit. The entire arrangement is located in a magnetic field $B = 0.15$ T directed out of the plane of the figure. The straight wire OP starts from rest with $\theta = 0$ and has a constant angular acceleration of 12 rad/s^2. (*a*) Find the resistance of the loop $OPQO$ as a function of θ. (*b*) Find the magnetic flux through the loop as a function of θ. (*c*) For what value of θ is the induced current in the loop a maximum? (*d*) What is the maximum value of the induced current in the loop?

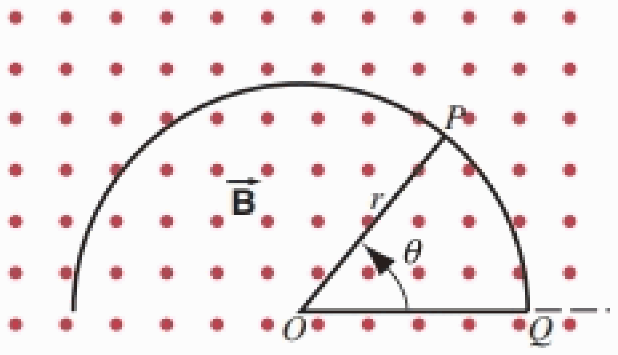


FIGURE 34-60. Problem 10.

11. An electromagnetic "eddy current" brake consists of a disk of conductivity σ and thickness t rotating about an axis through its center with a magnetic field $\vec{\mathbf{B}}$ applied perpendicular to the plane of the disk over a small area a^2 (see Fig. 34-61). If the area a^2 is at a distance r from the axis, find an approximate expression for the torque tending to slow down the disk at the instant its angular velocity equals ω.

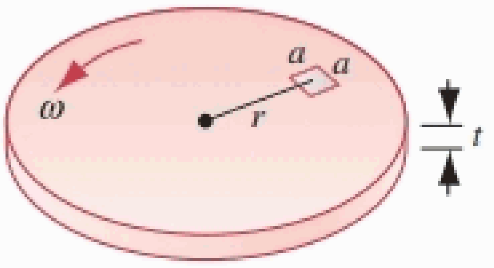


FIGURE 34-61. Problem 11.

12. A superconducting wire shaped in a ring of radius R originally carries no current. A magnet is moved along the axis of the ring, and the flux through the ring changes by an amount $\Delta\Phi_B$. Show that the current in the ring is given by

$$i = \frac{ne^2a^2}{2Rm_e}\,\Delta\Phi_B,$$

where $a \ll R$ is the radius of the wire, n the density of conducting electrons, and e and m_e the charge and mass of the electron, respectively.

13. Prove that the electric field $\vec{\mathbf{E}}$ in a charged parallel-plate capacitor cannot drop abruptly to zero as one moves at right angles to it, as suggested by the arrow in Fig. 34-62 (see point *a*). In actual capacitors fringing of the lines of force always occurs, which means that $\vec{\mathbf{E}}$ approaches zero in a continuous and gradual way; compare with Problem 11 of Chapter 33. (Hint: Apply Faraday's law to the rectangular path shown by the dashed lines.)

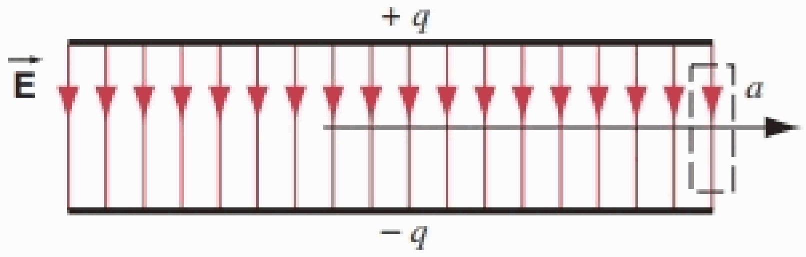


FIGURE 34-62. Problem 13.

14. A uniform magnetic field $\vec{\mathbf{B}}$ fills a cylindrical volume of radius R. A metal rod of length L is placed as shown in Fig. 34-63. If B is changing at the rate dB/dt, show that the emf that is produced by the changing magnetic field and that acts between the ends of the rod is given by

$$\mathcal{E} = \frac{dB}{dt}\,\frac{L}{2}\sqrt{R^2 - \left(\frac{L}{2}\right)^2}.$$

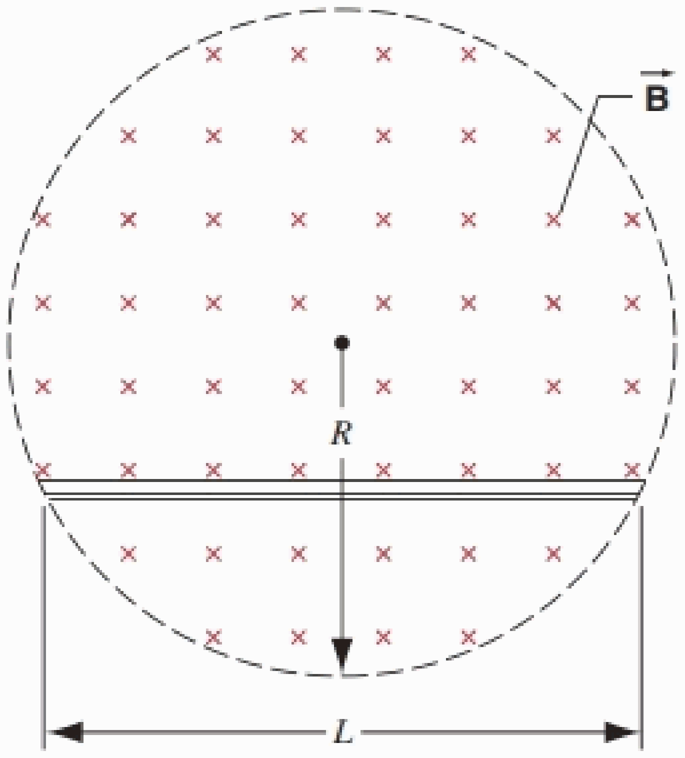


FIGURE 34-63. Problem 14.

15. In a certain betatron, the radius of the electron orbit is 32 cm and the magnetic field at the orbit is given by $B_{\text{orb}} = (0.28\ \text{T})\sin(120\pi\ \text{s}^{-1})t$. In the betatron, the average value B_{av} of the field enclosed by the electron orbit is equal to twice the value B_{orb} at the electron orbit. (*a*) Calculate the induced electric field felt by the electrons at $t = 0$. (*b*) Find the acceleration of the electrons at this instant. Ignore relativistic effects.

COMPUTER PROBLEM

1. Some measurements of the maximum magnetic field as a function of radius for a betatron are:

r (cm)	B (tesla)	r (cm)	B (tesla)
0	0.950	81.2	0.409
10.2	0.950	83.7	0.400
68.2	0.950	88.9	0.381
73.2	0.528	91.4	0.372
75.2	0.451	93.5	0.360
77.3	0.428	95.5	0.340

Show by graphical analysis that the relation $B_{av} = 2B_{orb}$ mentioned in Problem 15 as essential to betatron operation is satisfied at the orbit radius $R = 84$ cm. (Hint: Note that

$$B_{av} = \frac{1}{\pi R^2}\int_0^R B(r)2\pi r dr,$$

and evaluate the integral numerically.)

CHAPTER 35

MAGNETIC PROPERTIES OF MATERIALS

Magnetic materials play increasingly important roles in our daily lives. Materials such as iron, which can be permanent magnets at ordinary temperatures, are commonly used in electric motors and generators as well as in certain types of loudspeakers. Other materials can be "magnetized" and "demagnetized" with relative ease; these materials have found wide use for storing information in such applications as magnetic recording tape (used in audio tape recorders and VCRs), computer disks and credit cards. Still other materials are analogous to dielectrics in that they acquire an induced magnetic field in response to an external magnetic field; the induced field vanishes when the external field is removed.

In this chapter we consider the internal structure of materials that is responsible for their magnetic properties. We show that the behavior of different magnetic materials can be understood in terms of the magnetic dipole moments of individual atoms. A complete understanding of magnetic properties requires methods of quantum mechanics that are beyond the level of this text, but a qualitative understanding can be achieved based on principles discussed in this chapter. Finally, we consider a magnetic form of Gauss' law, which takes into account the apparent nonexistence of isolated magnetic poles.

35-1 THE MAGNETIC DIPOLE

For static electric fields, the single isolated charge is the fundamental quantity. Individual charges *produce* an electric field, and in turn the electric field set up by one group of charges can *influence* the behavior of other charges. On the basis of this elementary interaction between electric charges, we can explain many common phenomena: the force exerted by the nucleus on the electrons, which holds the atom together; the force exerted by one atom on another in ionic molecules and solids; elastic and frictional forces; and so forth.

In some electrically neutral molecules, it is useful to regard the fundamental interaction to be based on the electric dipole (which in turn can be analyzed as two point charges). We have seen how the dipole can *produce* an electric field (Section 26-3) and also how a dipole is *influenced* by other electric fields (Section 26-7).

For steady magnetic fields, the fundamental quantity is the moving electric charges in a current element, which can *produce* a magnetic field and can also be *influenced* by the magnetic field due to other current elements. However, in trying to explain the magnetic properties of materials, this explanation in terms of current elements is not as convenient as one based on the *magnetic dipole.* Ultimately the magnetic dipole can be regarded as caused by moving charges, just as the electric dipole can be regarded as two static charges. However, when we discuss the magnetic properties of materials we gain more insight by considering the materials to be a collection of atoms with individual magnetic dipole moments.

Let us begin by considering the magnetic field due to a circular current loop (Section 33-2) at a point on the z axis:

$$B = \frac{\mu_0 i R^2}{2(R^2 + z^2)^{3/2}} . \qquad (35\text{-}1)$$

If we are far from the loop ($z >> R$), this becomes

$$B = \frac{\mu_0 i R^2}{2z^3} = \frac{\mu_0}{2\pi}\frac{i\pi R^2}{z^3}. \tag{35-2}$$

The quantity $i\pi R^2$ in Eq. 35-2 can be written as iA, where $A = \pi R^2$ is the area of the circular loop. We define this quantity to be the magnitude of the *magnetic dipole moment* μ of the loop:

$$\mu = iA. \tag{35-3}$$

The magnetic dipole moment of a current loop is the product of the current and the area of the loop. Although we derived Eq. 35-3 for a circular loop, it holds for loops of any shape. If the loop has N turns, then $\mu = NiA$.

Equation 35-3 suggests that the units for μ are $\text{A}\cdot\text{m}^2$ (ampere-meter^2). Later in this section we will find that equivalent units are J/T (joules per tesla).

Like the electric dipole moment, the magnetic dipole moment is a vector quantity. The direction of $\vec{\boldsymbol{\mu}}$ is perpendicular to the plane of the current loop, determined by the right-hand rule: if the fingers of your right hand are in the direction of the current, then your thumb indicates the direction of $\vec{\boldsymbol{\mu}}$ (Fig. 35-1). With this definition, we can write Eq. 35-2 as a vector equation:

$$\vec{\mathbf{B}} = \frac{\mu_0 \vec{\boldsymbol{\mu}}}{2\pi z^3}. \tag{35-4}$$

Note that $\vec{\mathbf{B}}$ and $\vec{\boldsymbol{\mu}}$ are vectors in the same direction, as Fig. 35-1 shows. In Eq. 35-4, $\vec{\mathbf{B}}$ is the magnetic field *produced by* the magnetic moment $\vec{\boldsymbol{\mu}}$.

Let us now consider the effect of a magnetic field on a magnetic dipole. Figure 35-2 shows a current loop in a uniform magnetic field $\vec{\mathbf{B}}$. (This field $\vec{\mathbf{B}}$ is produced by some external agent not shown in the figure, perhaps a large solenoid.) In Section 32-6 we considered a similar problem (see Fig. 32-26) and concluded that in a uniform field the loop experiences no net force but does experience a net torque given by $\vec{\boldsymbol{\tau}} = iA\hat{\mathbf{n}} \times \vec{\mathbf{B}}$ (Eq. 32-35), where $\hat{\mathbf{n}}$ is a unit vector perpendicular to the loop in a direction determined by the right-hand rule. Because we defined the directions of $\vec{\boldsymbol{\mu}}$ and $\hat{\mathbf{n}}$ in exactly the same way, we can write $\vec{\boldsymbol{\mu}} = iA\hat{\mathbf{n}}$, and so Eq. 32-35 becomes

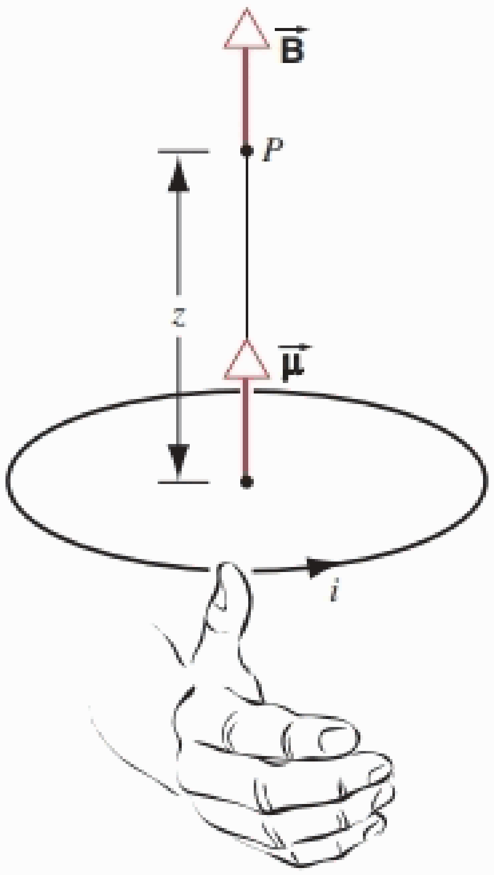


FIGURE 35-1. The magnetic dipole moment of a current loop and the magnetic field at point P a distance z from the loop on its axis.

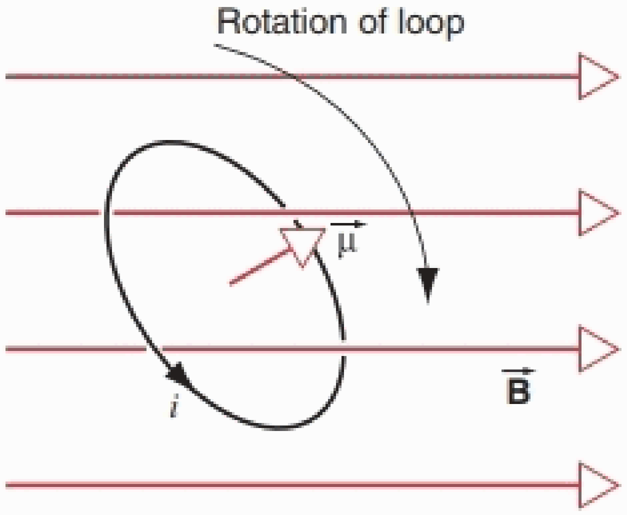


FIGURE 35-2. In an external magnetic field, a magnetic dipole moment experiences a torque that rotates $\vec{\boldsymbol{\mu}}$ into alignment with $\vec{\mathbf{B}}$.

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{\mu}} \times \vec{\mathbf{B}}. \tag{35-5}$$

That is, the torque tends to rotate the loop so that $\vec{\boldsymbol{\mu}}$ lines up with $\vec{\mathbf{B}}$. Note the similarity of Eq. 35-5 with the corresponding result for the torque that rotates an electric dipole in an electric field: $\vec{\boldsymbol{\tau}} = \vec{\mathbf{p}} \times \vec{\mathbf{E}}$ (Eq. 26-27). Equation 35-5 is valid regardless of the shape of the loop or its orientation relative to the magnetic field.

Equations 35-4 and 35-5 satisfy our two goals: Eq. 35-4 indicates how a magnetic field is *produced* by a magnetic dipole, and Eq. 35-5 shows how a magnetic dipole is *influenced by* an applied magnetic field. Keeping these two concepts in mind will help us understand the magnetic behavior of materials.

We can continue the analogy between electric and magnetic fields by considering the work done to change the orientation of a magnetic dipole in a magnetic field and relating that work to the potential energy of a magnetic dipole in a magnetic field. We can write the potential energy as

$$U = -\mu B\cos\theta = -\vec{\boldsymbol{\mu}}\cdot\vec{\mathbf{B}}, \tag{35-6}$$

for a magnetic dipole whose moment $\vec{\boldsymbol{\mu}}$ makes an angle θ with $\vec{\mathbf{B}}$. This equation is similar to the corresponding expression for an electric dipole, $U = -\vec{\mathbf{p}}\cdot\vec{\mathbf{E}}$ (Eq. 26-32). In Eq. 35-6, $U = 0$ when $\theta = 90°$ ($\vec{\boldsymbol{\mu}}$ is perpendicular to $\vec{\mathbf{B}}$ or equivalently $\vec{\mathbf{B}}$ is parallel to the plane of the loop). U has its smallest value ($= -\mu B$) when $\vec{\boldsymbol{\mu}}$ and $\vec{\mathbf{B}}$ are parallel, and U is largest $= +\mu B$) when $\vec{\boldsymbol{\mu}}$ and $\vec{\mathbf{B}}$ are antiparallel.

The magnetic force, like all forces that depend on velocity, is in general *not* conservative and therefore cannot generally be represented by a potential energy. In this special case, in which the torque on a dipole depends on its posi-

TABLE 35-1 **Selected Values of Magnetic Dipole Moments**

System	μ (J/T)
Nucleus of nitrogen atom	2.04×10^{-28}
Proton	1.41×10^{-26}
Neutron	9.65×10^{-27}
Electron	9.28×10^{-24}
Nitrogen atom	2.8×10^{-23}
Typical small coil[a]	5.4×10^{-6}
Small bar magnet	5
Superconducting coil	400
The Earth	8.0×10^{22}

[a] That of Sample Problem 35-1, for instance.

tion relative to the field, it *is* possible to define a potential energy for the *system* consisting of the dipole in the field. Note that the potential energy is not characteristic of the field alone, but of the dipole *in* the field. In general, we cannot define a scalar "magnetic potential energy" of a point charge or "magnetic potential" of the field itself such as we did for electric fields in Chapter 28.

Many physical systems have magnetic dipole moments: the Earth, bar magnets, current loops, atoms, nuclei, and elementary particles. Table 35-1 gives some typical values.

Note that Eq. 35-6 suggests units for μ of energy divided by magnetic field, or J/T. Equation 35-3 suggests units of current times area, or $\mathrm{A\cdot m^2}$. You can show that these two units are equivalent, and the choice between them is one of convenience.

As indicated by the examples of the proton and the nitrogen atom in Table 35-1, nuclear magnetic dipole moments are typically three to six orders of magnitude smaller than atomic magnetic dipole moments. Several conclusions follow immediately from this observation. (1) Electrons cannot be constituents of the nucleus; otherwise nuclear magnetic dipole moments would typically have magnitudes about the same as that of the electron. (2) Ordinary magnetic effects in materials are determined by *atomic* magnetism, rather than the much weaker *nuclear* magnetism. (3) To exert a particular torque necessary to align nuclear dipoles requires a magnetic field about three to six orders of magnitude larger than that necessary to align atomic dipoles.

SAMPLE PROBLEM 35-1. (*a*) A 250-turn rectangular coil of length 2.10 cm and width 1.25 cm carries a current of 85 μA. What is the magnetic dipole moment of this coil? (*b*) The magnetic dipole moment of the coil is lined up with an external magnetic field whose strength is 0.85 T. How much work would be done by an external agent to rotate the coil through 180°?

Solution (*a*) The magnitude of the magnetic dipole moment of the coil, whose area A is $(0.0210\ \mathrm{m})(0.0125\ \mathrm{m}) = 2.52 \times 10^{-4}\ \mathrm{m}^2$, is

$$\begin{aligned}\mu = NiA &= (250)(85 \times 10^{-6}\ \mathrm{A})(2.52 \times 10^{-4}\ \mathrm{m}^2)\\ &= 5.36 \times 10^{-6}\ \mathrm{A\cdot m^2} = 5.36 \times 10^{-6}\ \mathrm{J/T}.\end{aligned}$$

(*b*) The external work is equal to the increase in potential energy of the system, which is

$$\begin{aligned}W = \Delta U &= -\mu B\cos 180^\circ - (-\mu B\cos 0^\circ) = 2\mu B\\ &= 2(5.36 \times 10^{-6}\ \mathrm{J/T})(0.85\ \mathrm{T}) = 9.1 \times 10^{-6}\ \mathrm{J} = 9.1\ \mu\mathrm{J}.\end{aligned}$$

This is about equal to the work needed to lift an aspirin tablet through a vertical height of about 3 mm.

The Field of a Dipole

So far we have discussed the field of a magnetic dipole (a current loop) only at points on the axis. Now we consider the complete magnetic dipole field. In the case of the electric dipole, a complete pattern of field lines was shown in Fig. 26-12. A few field lines for an electric dipole are shown in Fig. 35-3*a* and can be compared with the field lines for a current loop shown in Fig. 35-3*b*. You can see a great similarity between the pattern of field lines outside the loop. Another similarity between the electric and magnetic dipole fields is that both vary as r^{-3} when we are far from the dipole. A significant difference between electric and magnetic field lines is that electric field lines start on positive charges and end on negative charges, whereas magnetic field lines always form closed loops.

Figure 35-3*c* shows the field lines of a bar magnet. It shows the same pattern of field lines as the current loop, so a bar magnet can also be considered to be a magnetic dipole. It is convenient to label the two ends of a bar magnet as the north (N) and south (S) poles, with field lines leaving the N pole and converging on the S pole. Superficially the poles may seem to behave like the positive and negative charges of an electric dipole. However, close inspection of Fig. 35-3*c* shows that the field lines do not start and end on the poles but instead continue through the interior of the magnet, again forming closed loops. The N and S poles do not behave like the charges in an electric dipole, and as we discuss in Section 35-7, isolated magnetic poles do not appear to exist in nature.

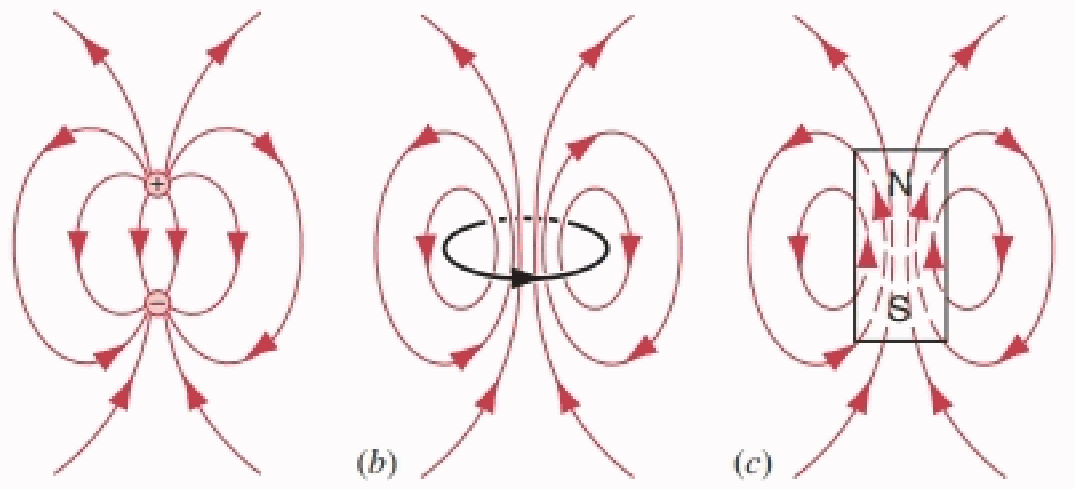


FIGURE 35-3. (*a*) The electric field of an electric dipole. (*b*) The dipole magnetic field of a current loop. (*c*) The dipole magnetic field of a bar magnet. The dashed lines show the field lines inside the magnet.

35-2 THE FORCE ON A DIPOLE IN A NONUNIFORM FIELD

In a uniform electric field, the forces on the two charges of an electric dipole are of equal magnitude and in opposite directions (Fig. 26-19). If the field is nonuniform, the forces are of different magnitudes and so a net force can act on the dipole. The same conclusion holds for magnetic dipoles: in a uniform magnetic field, there may be a net torque on the dipole, but there is no net force. For a net force to be exerted on the dipole, the magnetic field must be nonuniform.

Consider the pair of current loops shown in Fig. 35-4. The loops lie along a common axis, and both loops carry current in the counterclockwise direction as viewed from above. Loop 1 sets up a magnetic field $\vec{\mathbf{B}}_1$, which then interacts with loop 2. (We assume that loop 2 has already been rotated by the torque due to the field of loop 1, which lines up the dipole moment of loop 2 with the field of loop 1.) At points C and D, which are on opposite ends of a diameter of loop 2, the forces $d\vec{\mathbf{F}} = i_2\, d\vec{\mathbf{s}} \times \vec{\mathbf{B}}_1$ on the elements $d\vec{\mathbf{s}}$ have downward and radially outward components. When we add the forces on all such pairs of elements, we find that the radial components cancel and the downward components sum to give a net downward force on the current loop.

We can also analyze this force in terms of magnetic poles. Each of the current loops can be represented as a magnet with north and south poles oriented as shown in Fig. 35-4. The attraction of loop 1 for loop 2 can be described in terms of the force between magnetic poles: the N pole of the magnet representing loop 1 attracts the S pole of the magnet representing loop 2. In Fig. 35-4, there is also a repulsion between the two N poles and the two S poles, but the N–S attraction is the stronger force because the poles are closer together.

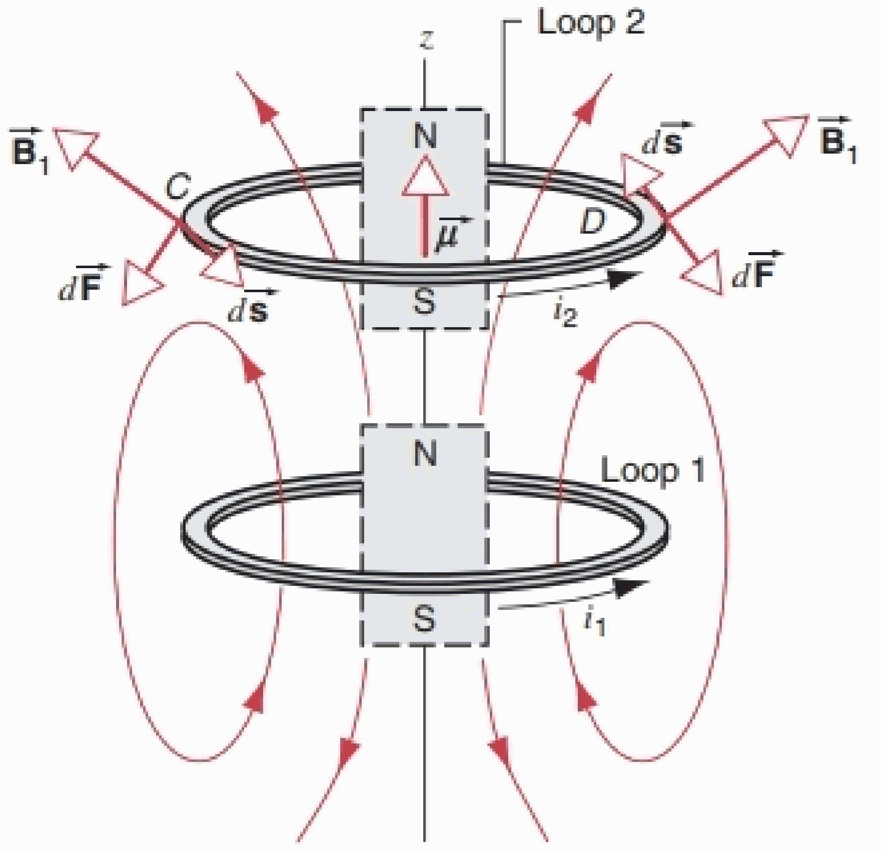


FIGURE 35-4. The magnetic field $\vec{\mathbf{B}}_1$ due to loop 1 causes a net downward force on loop 2.

Using Eq. 35-6 for the potential energy of the magnetic dipole moment of loop 2 in the magnetic field caused by loop 1 ($U = -\vec{\boldsymbol{\mu}}_2 \cdot \vec{\mathbf{B}}_1$), we obtain $U = -\mu_{2z}B_{1z}$, because the magnetic moment of loop 2 has only a z component. The z component of the force $\vec{\mathbf{F}}_{21}$ exerted on loop 2 by loop 1 is related to the potential energy by $F_z = -dU/dz$, so

$$F_{21z} = -\frac{dU}{dz} = -\frac{d}{dz}(-\mu_{2z}B_{1z}) = \mu_{2z}\frac{dB_{1z}}{dz}. \quad (35\text{-}7)$$

In Fig. 35-4, taking the z axis as positive upward, we have $\mu_{2z} > 0$ and $dB_{1z}/dz < 0$ (since the z component of the field decreases as we go upward), and so $F_{21z} < 0$. The force on loop 2 due to loop 1 is downward, as we have already determined.

Similar considerations show that the force on loop 1 due to loop 2 is upward, so the two loops attract each other.

Induced Magnetic Dipole Moments

In some materials in which the molecules do not have a permanent electric dipole moment, as we discussed in Section 29-6, an applied electric field can induce a dipole moment by causing a separation of the positive and negative charges in the molecule. A similar effect occurs for magnetic fields: in materials that lack permanent magnetic dipole moments, an applied magnetic field can induce a dipole moment.

Figure 35-5 shows how this might occur. Consider a double loop, consisting of two single loops carrying identical currents in opposite directions, in a nonuniform field that might be produced by a permanent magnet. The net magnetic moment of the double loop is zero, because the two single loops have magnetic moments of equal magni-

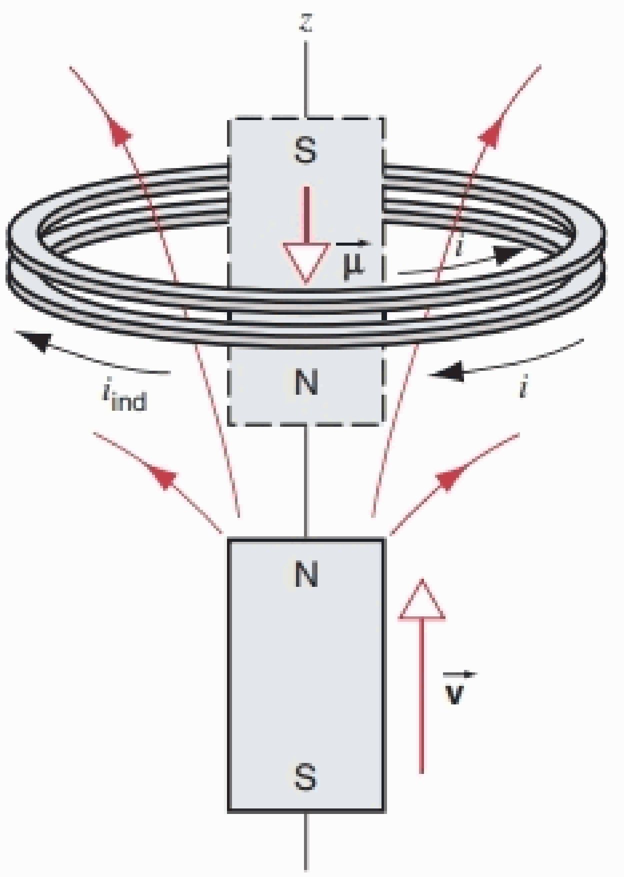


FIGURE 35-5. The double loop has no permanent magnetic dipole moment, but it acquires an induced dipole moment when the magnet approaches the loop. The loop is repelled by the force on the induced moment.

tudes but opposite directions. As the magnet is brought closer to the double loop, the flux through the loops increases, causing an induced current that, according to Lenz' law, must be directed clockwise (viewed from above). This induced current, which adds to the currents in the two loops, gives a net current $i - i_{\text{ind}}$ in the upper loop and $i + i_{\text{ind}}$ in the lower loop. The result is a net induced magnetic moment directed downward. The N and S poles of the equivalent magnet are shown, from which it can be seen that the force on the double loop due to the magnet is repulsive (upward).

Using Eq. 35-7, we see that (again taking the z axis as positive upward), $\mu_z < 0$ and $dB_z/dz < 0$, so that $F_z > 0$, corresponding to an upward force, in agreement with the previous conclusion.

In summary, in a nonuniform magnetic field, permanent dipoles are rotated into alignment with the field and attracted to the source of the field, but induced dipoles are repelled from the source of the field.

35-3 ATOMIC AND NUCLEAR MAGNETISM

The bulk electrical properties of a dielectric substance such as water depend on the individual electric dipole moments of its molecules. Each molecule has a positive side and a negative side and behaves like an electric dipole. If we break apart the molecule we can separate the positive and negative parts.

The magnetic properties of materials similarly depend on the magnetic dipole moments of individual atoms, and we can consider magnetic materials to be composed of a collection of atomic dipoles, which might align when an external magnetic field is applied (Fig. 35-6). However, unlike the electric dipole, we cannot break apart the atoms into separate N and S magnetic poles. Instead, we consider the magnetic moments to be tiny current loops, caused for example by the circulation of electrons in orbits in the atom. In this section we discuss the magnetic dipole moment associated with a circulating electron.

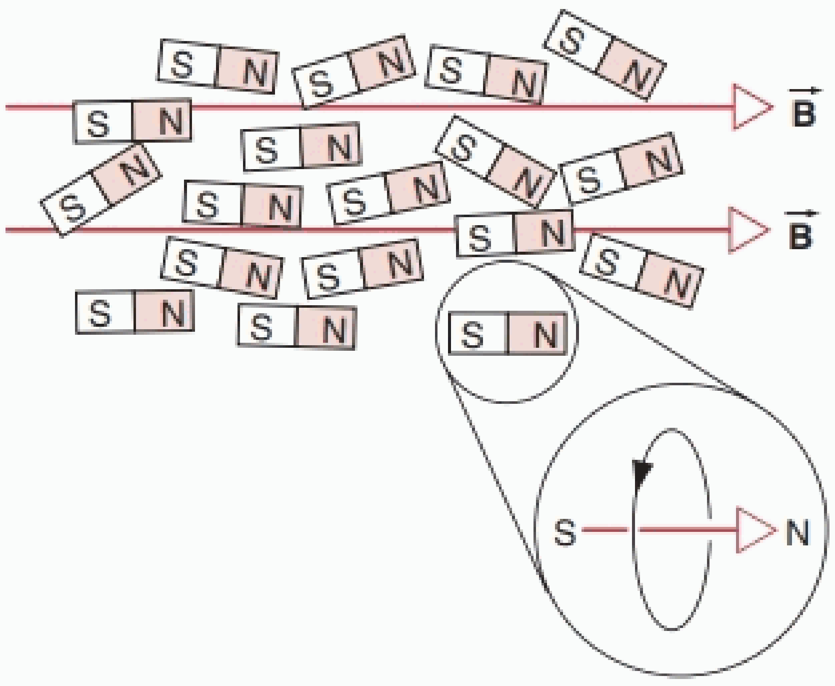


FIGURE 35-6. A magnetic material can be regarded as a collection of magnetic dipole moments, each with a north and a south pole. Microscopically, each dipole is actually a current loop that cannot be split into individual poles.

We consider a simple model of an atom in which an electron moves in a circular orbit of radius r and speed v about the nucleus. This circulating electron can be considered as a current loop, in which the current is the magnitude of the charge of the electron divided by the period T for one orbit:

$$i = \frac{e}{T} = \frac{e}{2\pi r/v}. \tag{35-8}$$

The magnetic dipole moment of the loop can be found using Eq. 35-3:

$$\mu = iA = \left(\frac{ev}{2\pi r}\right)(\pi r^2) = \frac{erv}{2}. \tag{35-9}$$

The magnetic dipole moment we have computed here for atoms is known as the *orbital* magnetic dipole moment, because it is due to the orbital motion of electrons about the nucleus.

In analyzing the properties of atoms, it is convenient to rewrite Eq. 35-9 as

$$\mu_l = \frac{erv}{2} = \frac{e}{2m}mvr = \frac{e}{2m}l, \tag{35-10}$$

where m is the mass of an electron. The quantity mvr is the angular momentum l of the electron moving in a circular orbit about the nucleus of the atom. We have labeled the orbital magnetic dipole moment as μ_l to indicate that it arises from the orbital angular momentum l. In the quantum theory of atoms, which is discussed in Chapters 47 and 48, the angular momentum is measured in units of $h/2\pi$, where h is the Planck constant. Substituting this fundamental unit of angular momentum into Eq. 35-10, we obtain a basic unit of the magnetic dipole moment called the *Bohr magneton* μ_{B}:

$$\mu_{\text{B}} = \frac{e}{2m}\frac{h}{2\pi} = \frac{eh}{4\pi m} = 9.27 \times 10^{-24}\ \text{J/T}, \tag{35-11}$$

where the numerical value is obtained by inserting the numerical values of e, h, and m into Eq. 35-11. Atomic magnetic moments are usually measured in units of μ_{B} and are typically on the order of 1 μ_{B} in magnitude, as you can see from the example of the nitrogen atom in Table 35-1.

The magnetic dipole moments of atoms can be measured by passing a beam of atoms through a region in which there is a nonuniform magnetic field. As we showed in the previous section, there is a net force on a magnetic dipole in a nonuniform field, so the atoms are deflected from their original paths when they pass through the field region. In the 1920s, experiments of this type showed that atoms with no orbital magnetic dipole moments were still deflected by a magnetic field. This result suggests the presence of another contribution to the magnetic dipole moments of atoms, in this case coming from the magnetic dipole moments of the electrons themselves, called the *intrinsic* or "spin" magnetic moment. Electrons in different states of motion have different orbital magnetic dipole moments, but all electrons have exactly the same intrinsic magnetic dipole moment. The intrinsic spin of the electron and its effect on the structure of atoms is discussed in Chapters 47 and 48. The intrinsic magnetic dipole moment of an electron is listed in Table 35-1. Its value is very nearly equal to 1 μ_B.

The orbital and spin magnetic dipole moments of electrons are about the same size (of order 1 μ_B), and thus both are important in determining the magnetic properties of atoms. The total magnetic moment of an atom is obtained from the vector sum of the orbital and spin magnetic moments of all its electrons. In an atom with many electrons, these vector sums can be very complicated. In some atoms, however, the total orbital and spin magnetic moments turn out to be zero. Materials made from these atoms are virtually nonmagnetic and show only a weak induced effect called *diamagnetism* (analogous to that of Fig. 35-5). In other atoms, the total orbital or spin magnetic moment (or both) may be nonzero, so that the atoms align in the presence of an external magnetic field. These materials are called *paramagnetic.* The most familiar type of magnetic behavior is *ferromagnetism,* in which, because of the interactions among the atoms, the alignment of the atoms remains even when the external field is removed. Later in this chapter we discuss these three kinds of magnetic materials in more detail.

Nuclear Magnetism

The nucleus, which is composed of protons and neutrons in orbital motion under the influence of their mutual forces, has a magnetic moment with two parts: an orbital part, due to the motion of the protons (neutrons, being uncharged, do not contribute to the orbital magnetic moment even though they may have orbital angular momentum), and an intrinsic part, due to the intrinsic magnetic moments of the protons and neutrons, which are listed in Table 35-1. (It may seem surprising that the uncharged neutron has a nonzero intrinsic magnetic moment. If the neutron were truly an elementary particle with no electric charge, it would indeed have no magnetic dipole moment. The nonzero magnetic dipole moment of the neutron is a clue to its internal structure and can be fairly well accounted for by considering the neutron to be composed of three charged quarks.)

Nuclei have orbital and spin magnetic dipole moments that can be expressed in the form of Eq. 35-10. However, the mass that appears in these equations (the electron mass) must be replaced by the proton or neutron mass, which is about 1800 times the electron mass. Typical nuclear magnetic dipole moments are smaller than atomic dipole moments by a factor of the order of 10^{-3} (see Table 35-1), and their contribution to the magnetic properties of materials is usually negligible.

The effects of nuclear magnetism become important in the case of *nuclear magnetic resonance,* in which the nucleus is subject to electromagnetic radiation of a precisely defined frequency corresponding to that necessary to cause the nuclear magnetic moment to change direction. We can align the nuclear magnetic moments in a sample of material by a static magnetic field; the direction of the dipoles reverses when they absorb the time-varying electromagnetic radiation. The absorption of this radiation can easily be detected. This effect is the basis of magnetic resonance imaging (MRI), a diagnostic technique in which images of organs of the body can be obtained using radiation far less dangerous to the body than x rays (Fig. 35-7).

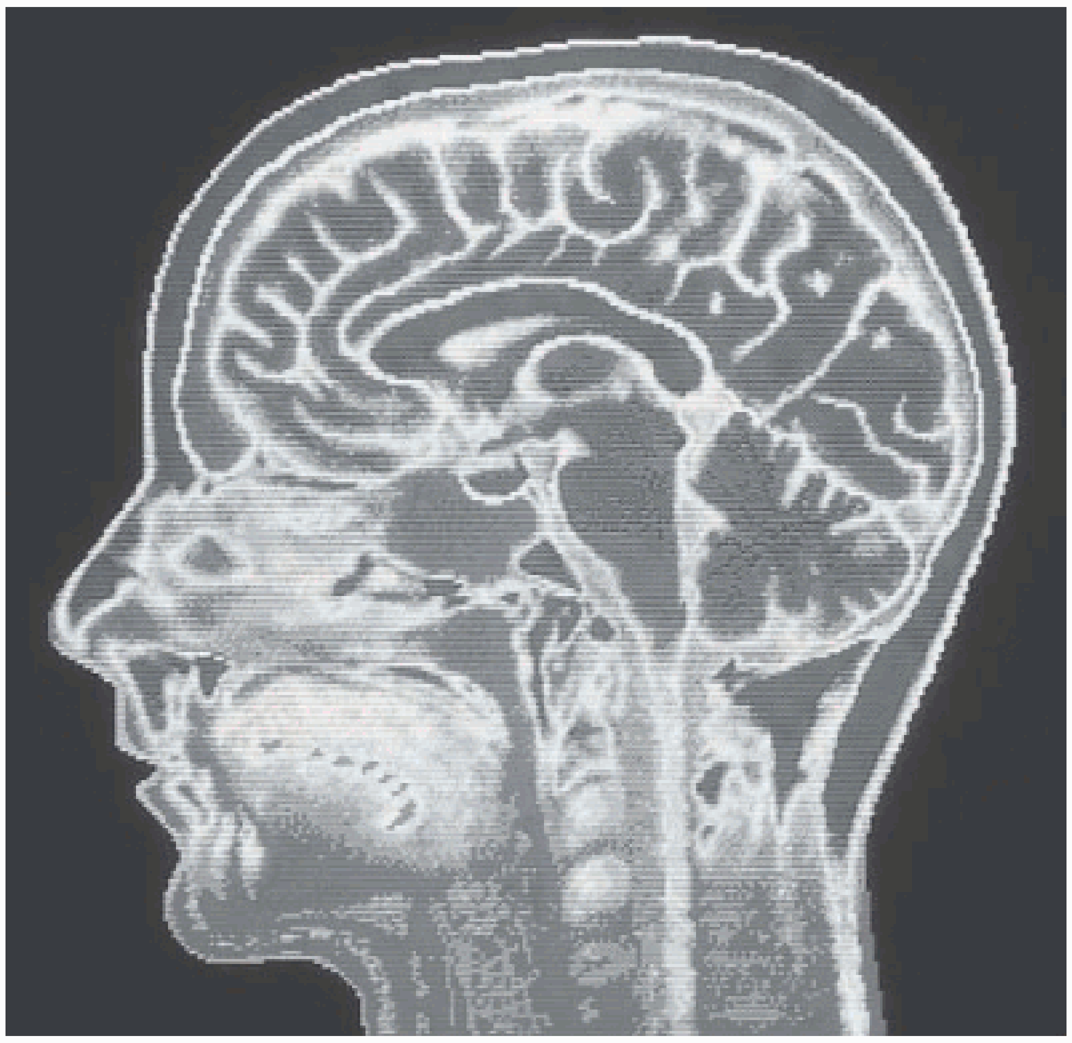


FIGURE 35-7. A profile of a human head, taken by magnetic resonance imaging (MRI) techniques. It shows details of the brain and facial tissue not visible on x-ray images and involves no radiation health risk to the patient.

SAMPLE PROBLEM 35-2. A proton is in a magnetic field of strength $B = 1.5$ T. The magnetic dipole moment of the proton is initially antiparallel to the direction of $\vec{\mathbf{B}}$. How much external work must be done to reverse the direction of the magnetic dipole moment of the proton?

Solution The energy of interaction of a magnetic dipole with a magnetic field was given by Eq. 35-6, $U = -\vec{\mu}\cdot\vec{\mathbf{B}}$. When $\vec{\mu}$ is antiparallel to the field, as in the initial state of this problem, the initial energy U_i is

$$U_i = -\vec{\mu}\cdot\vec{\mathbf{B}} = \mu B,$$

because the angle between $\vec{\mu}$ and $\vec{\mathbf{B}}$ is 180°. When the magnetic dipole moment changes direction (called a "spin flip"), the magnetic moment becomes *parallel* to $\vec{\mathbf{B}}$, and the final energy is

$$U_f = -\vec{\mu}\cdot\vec{\mathbf{B}} = -\mu B.$$

The external work done on the system is equal to the change in energy, or

$$\begin{aligned} W = U_f - U_i &= -\mu B - \mu B = -2\mu B \\ &= -2(1.41 \times 10^{-26}\ \mathrm{J/T})(1.5\ \mathrm{T}) \\ &= -4.23 \times 10^{-26}\ \mathrm{J} = -0.26\ \mu\mathrm{eV}. \end{aligned}$$

Because the environment does *negative* work on the system, the system does *positive* work on its environment. This energy might be transmitted to the environment in the form of electromagnetic radiation, which would be in the radio-frequency range of the spectrum and would have a frequency of 64 MHz, slightly below the tuning range of an FM radio.

35-4 MAGNETIZATION

In Chapter 30 we considered the effect of filling the space between capacitor plates with a dielectric medium, and we found that inserting the dielectric while keeping the charge on the plates constant reduced the electric field in the region between the plates. That is, if $\vec{\mathbf{E}}_0$ is the electric field without the dielectric, then the field $\vec{\mathbf{E}}$ with the dielectric is given by Eq. 29-23, which we write in vector form as

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0/\kappa_e. \qquad (35\text{-}12)$$

The effect of the dielectric is characterized by the dielectric constant κ_e, a pure number with a value greater than 1 for materials (see Table 29-2).

Consider instead a magnetic medium composed of atoms having magnetic dipole moments $\vec{\boldsymbol{\mu}}_n$. These dipoles in general point in various directions in space. Let us compute the net dipole moment $\vec{\boldsymbol{\mu}}$ of a volume V of the material by taking the *vector* sum of all the dipoles in that volume: $\vec{\boldsymbol{\mu}} = \Sigma\vec{\boldsymbol{\mu}}_n$. We then define the *magnetization* $\vec{\mathbf{M}}$ of the medium to be the net dipole moment per unit volume, or

$$\vec{\mathbf{M}} = \frac{\vec{\boldsymbol{\mu}}}{V} = \frac{\Sigma\vec{\boldsymbol{\mu}}_n}{V}. \qquad (35\text{-}13)$$

For the magnetization to be considered a microscopic quantity, Eq. 35-13 should be written as the limit as the volume approaches zero. This permits us to consider a material as having a *uniform* magnetization.

Suppose such a material is placed in a uniform field $\vec{\mathbf{B}}_0$. This applied field "magnetizes" the material and aligns the dipoles. The aligned dipoles produce a magnetic field of their own, in analogy with the *electric* field produced by the electric dipoles in a dielectric medium (see Section 29-6). At any point in space, the net magnetic field $\vec{\mathbf{B}}$ is the sum of the applied field $\vec{\mathbf{B}}_0$ and that produced by the dipoles, which we call $\vec{\mathbf{B}}_M$, so that

$$\vec{\mathbf{B}} = \vec{\mathbf{B}}_0 + \vec{\mathbf{B}}_M. \qquad (35\text{-}14)$$

The field $\vec{\mathbf{B}}_M$ can include contributions from permanent dipoles in paramagnetic materials (analogous to polar dielectrics) and induced dipoles in all materials (as in nonpolar dielectrics).

The magnetization field $\vec{\mathbf{B}}_M$ is related to the magnetization $\vec{\mathbf{M}}$, which (as defined in Eq. 35-13) is also determined by the dipoles in the material. In weak fields, $\vec{\mathbf{M}}$ is proportional to the applied field $\vec{\mathbf{B}}_0$. However, $\vec{\mathbf{B}}_M$ is in general difficult to calculate unless the magnetization is uniform and the geometry has a high degree of symmetry. As an example of such a case, we consider a long (ideal) solenoid of circular cross section filled with a magnetic material (Fig. 35-8). In this case, the applied field is uniform throughout the interior; both $\vec{\mathbf{B}}_0$ and $\vec{\mathbf{M}}$ are parallel to the axis, and it can be shown that $\vec{\mathbf{B}}_M = \mu_0\vec{\mathbf{M}}$ in the interior of the solenoid. (You should check the dimensions and show that $\mu_0\vec{\mathbf{M}}$ has the same dimensions as $\vec{\mathbf{B}}$.)

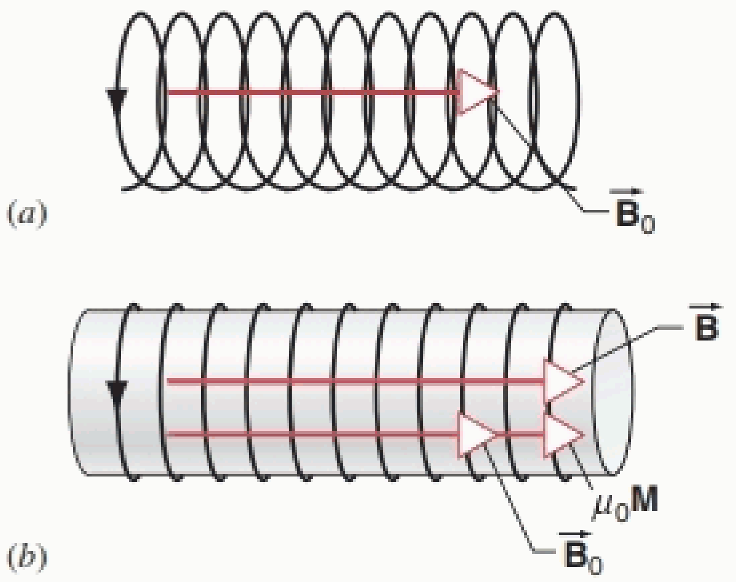


FIGURE 35-8. (*a*) In an empty solenoid, the current establishes a field $\vec{\mathbf{B}}_0$. (*b*) When the solenoid is filled with magnetic material, the total field $\vec{\mathbf{B}}$ includes contributions $\vec{\mathbf{B}}_0$ from the current and $\mu_0\vec{\mathbf{M}}$ from the magnetic material.

We can therefore write the net field as

$$\vec{\mathbf{B}} = \vec{\mathbf{B}}_0 + \mu_0\vec{\mathbf{M}}, \qquad (35\text{-}15)$$

as illustrated in Fig. 35-8*b*. In weak fields, $\vec{\mathbf{M}}$ increases linearly with the applied field $\vec{\mathbf{B}}_0$, and so $\vec{\mathbf{B}}$ must be proportional to $\vec{\mathbf{B}}_0$. In this case, we can write

$$\vec{\mathbf{B}} = \kappa_m\vec{\mathbf{B}}_0, \qquad (35\text{-}16)$$

where κ_m is the *permeability* of the material, which is defined relative to a vacuum, for which $\kappa_m = 1$. Permeabilities of most common materials (excepting ferromagnets) have values very close to 1, as we discuss in the next section. For materials other than ferromagnets, the permeability may depend on such properties as the temperature and density of the material, but not on the field $\vec{\mathbf{B}}_0$. Under ordinary circumstances, Eq. 35-16 describes a linear relationship with the net field $\vec{\mathbf{B}}$ increasing linearly as the applied field increases. For ferromagnets, on the other hand, we can regard Eq. 35-16 as defining a particular κ_m that depends on the applied field $\vec{\mathbf{B}}_0$, so that Eq. 35-16 is no longer linear.*

Combining Eqs. 35-15 and 35-16, we can write the magnetization induced by the applied field as

$$\mu_0\vec{\mathbf{M}} = (\kappa_m - 1)\,\vec{\mathbf{B}}_0. \qquad (35\text{-}17)$$

The quantity $\kappa_m - 1$ is typically of order 10^{-3} to 10^{-6} for most nonferromagnetic materials, and so the contribution of

---

*There is, as always, an analogy here between electric and magnetic fields. There are dielectric materials, called *ferroelectrics,* in which the relationship between $\vec{\mathbf{E}}$ and $\vec{\mathbf{E}}_0$ is nonlinear; that is, κ_e is dependent on the applied field $\vec{\mathbf{E}}_0$. From such materials we can construct quasipermanent electric dipoles, called *electrets,* which are analogous to permanent magnets. Most dielectric materials in common use are linear, whereas the most commonly useful magnetic materials are nonlinear.

the magnetization $\mu_0\vec{\mathbf{M}}$ to the total field is generally far smaller than $\vec{\mathbf{B}}_0$. This is in great contrast to the case of electric fields, in which κ_e has values for typical materials in the range of 3–100. The net electric field is modified substantially by the dielectric medium, but the magnetic medium has only a very small effect on the magnetic field for nonferromagnets.

SAMPLE PROBLEM 35-3. The magnetic field in the interior of a certain solenoid has the value 6.5×10^{-4} T when the solenoid is empty. When it is filled with iron, the field becomes 1.4 T. (*a*) Find the relative permeability under these conditions. (*b*) Find the average magnetic moment of an iron atom under these conditions.

Solution (*a*) From Eq. 35-16, we have (taking magnitudes only)

$$\kappa_m = \frac{B}{B_0} = \frac{1.4\ \mathrm{T}}{6.5 \times 10^{-4}\ \mathrm{T}} = 2200.$$

(*b*) Using Eq. 35-15, we obtain

$$M = \frac{B - B_0}{\mu_0} = \frac{1.4\ \mathrm{T} - 6.5 \times 10^{-4}\ \mathrm{T}}{4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}} = 1.11 \times 10^6\ \mathrm{A/m}.$$

Note that the units of M can also be expressed as $\mathrm{A \cdot m^2/m^3}$. This represents the magnetic moment per unit volume of the iron. To find the magnetic moment per atom, we need the number density n of atoms (the number of atoms per unit volume):

$$n = \frac{\text{atoms}}{\text{volume}} = \frac{\text{mass}}{\text{volume}}\,\frac{\text{atoms}}{\text{mass}}$$

$$= \frac{\text{mass}}{\text{volume}}\,\frac{\text{atoms/mole}}{\text{mass/mole}} = \rho\frac{N_A}{m}.$$

Here ρ is the density of iron, N_A is the Avogadro constant, and m is the molar mass of iron. Putting in the values, we obtain

$$n = (7.87 \times 10^3\ \mathrm{kg/m^3})\frac{6.02 \times 10^{23}\ \text{atoms/mol}}{0.0558\ \mathrm{kg/mol}}$$

$$= 8.49 \times 10^{28}\ \text{atoms/m}^3.$$

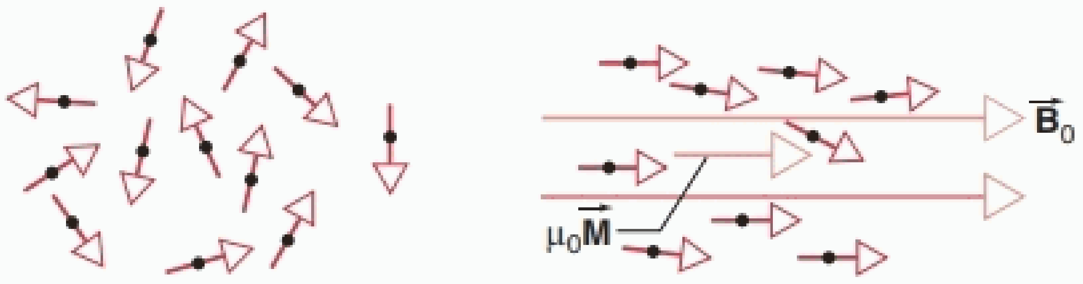
The average magnetic moment per atom is

$$\mu = \frac{M}{n} = \frac{1.11 \times 10^6\ \mathrm{A/m}}{8.49 \times 10^{28}/\mathrm{m^3}} = 1.31 \times 10^{-23}\ \mathrm{J/T} = 1.4\ \mu_B.$$

This result is quite consistent with what we expect for an atomic magnetic moment. The calculation suggests that each atom of the sample of iron is contributing its full magnetic dipole moment to the magnetization of the material, a situation that characterizes ferromagnets.

35-5 MAGNETIC MATERIALS

We are now ready to consider some characteristics of three types of magnetic materials. As we shall see, these classifications depend in part on the magnetic dipole moments of the atoms of the material and in part on the interactions among the atoms.



(*a*) (*b*)

FIGURE 35-9. (*a*) In an unmagnetized sample, the atomic magnetic moments are randomly oriented. (*b*) When an external field $\vec{\mathbf{B}}_0$ is applied, the dipoles rotate into alignment with the field, and the vector sum of the atomic dipole moments gives a contribution $\mu_0\vec{\mathbf{M}}$ to the field in the material.

Paramagnetism

Paramagnetism occurs in materials whose atoms have permanent magnetic dipole moments; it makes no difference whether these dipole moments are of the orbital or spin types.

In a sample of a paramagnetic material to which no magnetic field is applied, the atomic dipole moments initially are randomly oriented in space (Fig. 35-9*a*). The magnetization, computed according to Eq. 35-13, is zero, because the random directions of the $\vec{\boldsymbol{\mu}}_n$ cause the vector sum to vanish, just as the randomly directed velocities of the molecules in a sample of a gas sum to give zero for the center-of-mass velocity of the entire sample.

When an external magnetic field is applied to the material (perhaps by placing it within the windings of a solenoid), the resulting torque on the dipoles tends to align them with the field (Fig. 35-9*b*). The vector sum of the individual dipole moments is no longer zero. The field inside the material now has two components: the applied field $\vec{\mathbf{B}}_0$ and the induced field $\mu_0\vec{\mathbf{M}}$ from the magnetization of the dipoles. Note that these two fields are parallel; the dipoles enhance the applied field, in contrast to the electrical case in which the dipole field opposed the applied field and reduced the total electric field in the material (see Fig. 29-11). The ratio of $\mu_0\vec{\mathbf{M}}$ to $\vec{\mathbf{B}}_0$ is determined, according to Eq. 35-17, by $\kappa_m - 1$, which is small and positive for paramagnetic materials. Table 35-2 shows some representative values.

TABLE 35-2 **Relative Permeability of Some Paramagnetic Materials at Room Temperature**

Material	$\kappa_m - 1$
Gd_2O_3	1.2×10^{-2}
$CuCl_2$	3.5×10^{-4}
Chromium	3.3×10^{-4}
Tungsten	6.8×10^{-5}
Aluminum	2.2×10^{-5}
Magnesium	1.2×10^{-5}
Oxygen (1 atm)	1.9×10^{-6}
Air (1 atm)	3.6×10^{-7}

The thermal motion of the atoms tends to disturb the alignment of the dipoles, and consequently the magnetization decreases with increasing temperature. The relationship between M and the temperature T was discovered to be an inverse one by Pierre Curie in 1895 and is written

$$M = C\frac{B_0}{T}, \tag{35-18}$$

which is known as *Curie's law,* the constant C being known as the Curie constant. The temperature in Eq. 35-18 must be in units of kelvins. Equation 35-18 is valid only when B_0/T is small—that is, for small fields or high temperatures.

In large applied fields, the magnetization approaches its maximum value, which occurs when all the dipoles are parallel. If there are N such dipoles in the volume V, the maximum value of $\Sigma\vec{\mu}_n$ is $N\mu_n$, corresponding to the N parallel $\vec{\mu}_n$ vectors. In this case Eq. 35-13 gives

$$M_{\max} = \frac{N}{V}\mu_n. \tag{35-19}$$

When the magnetization reaches this *saturation* value, increases in the applied field $\vec{\mathbf{B}}_0$ have no further effect on the magnetization. Curie's law, which requires that M increase linearly with B_0, is valid only when the magnetization is far from saturation—that is, when B_0/T is small. Figure 35-10 shows the measured magnetization M, as a fraction of the maximum value $M_{\max}$, as a function of B_0/T for various temperatures for the paramagnetic salt chrome alum, $CrK(SO_4)_2 \cdot 12H_2O$. (It is the chromium ions in this salt that are responsible for the paramagnetism.) Note the approach to saturation, and note that Curie's law is valid only at small values of B_0/T (corresponding to small applied fields or high temperatures).

When the external magnetic field is removed from a paramagnetic sample, the thermal motion causes the directions of the magnetic dipole moments to become random again; the magnetic forces between atoms are too weak to hold the alignment and prevent the randomization. This effect can be used to achieve cooling in a process known as *adiabatic demagnetization.* A sample is magnetized at constant temperature. The dipoles move into a state of minimum energy in full or partial alignment with the applied field, and in doing so they must give up energy to the surrounding material. This energy flows as heat to the thermal reservoir of the environment. Now the sample is thermally isolated from its environment and is demagnetized adiabatically. When the dipoles become randomized, the increase in their magnetic energy must be compensated by a corresponding decrease in the internal energy of the system (since heat cannot flow to or from the isolated system in an adiabatic process). The temperature of the sample must therefore decrease. The lowest temperature that can be reached is determined by the residual field caused by the dipoles. The demagnetization of atomic magnetic dipoles can be used to achieve temperatures on the order of 0.001 K, whereas the demagnetization of the much smaller nuclear magnetic dipoles permits temperatures below 10^{-6} K to be obtained.

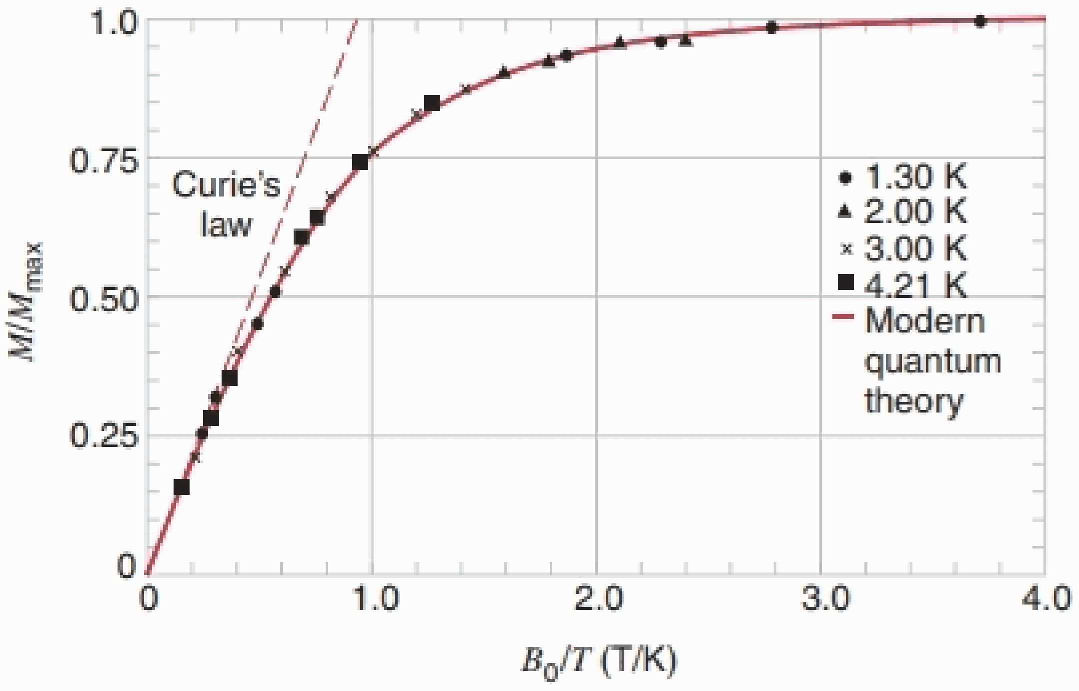


FIGURE 35-10. For a paramagnetic material, the ratio of the magnetization M to its saturation value $M_{\max}$ varies with B_0/T.

Diamagnetism

In 1847, Michael Faraday discovered that a specimen of bismuth was *repelled* by a strong magnet. He called such substances diamagnetic. (In contrast, paramagnetic substances are always *attracted* by a magnet.) Diamagnetism occurs in all materials. However, it is generally a much weaker effect than paramagnetism, and therefore it can most easily be observed only in materials that are not paramagnetic. Such materials might be those having atomic magnetic dipole moments of zero, perhaps originating from atoms having several electrons with their orbital and spin magnetic moments adding vectorially to zero.

Diamagnetism is analogous to the effect of induced electric fields in electrostatics. An uncharged bit of material such as paper is attracted to a charged rod of either polarity. The molecules of the paper do not have permanent electric dipole moments but acquire *induced* dipole moments from the action of the electric field, and these induced moments can then be attracted by the field (see Fig. 25-5).

In diamagnetic materials, atoms having no *permanent* magnetic dipole moments acquire *induced* dipole moments when they are placed in an external magnetic field, as we discussed in Section 35-2. Consider the orbiting electrons in an atom to behave like current loops. When an external field $\vec{\mathbf{B}}_0$ is applied, the flux through the loop changes. By Lenz' law, the motion must change in such a way that an induced field opposes this increase in flux. A calculation based on circular orbits (see Problem 7) shows that the change in motion is accomplished by a slight speeding up or slowing down of the orbital motion, such that the circular frequency associated with the orbital motion changes by

$$\Delta\omega = \pm\frac{eB_0}{2m}, \tag{35-20}$$

where B_0 is the magnitude of the applied field and m is the mass of an electron. This change in the orbital frequency in effect changes the orbital magnetic moment of an electron (see Eq. 35-8 and Sample Problem 35-5).

If we were to bring a single atom of a material such as bismuth near the north pole of a magnet, the field (which points away from the pole) would tend to increase the flux through the current loop that represents a circulating electron in the bismuth atom. According to Lenz' law, there must be an induced field pointing in the opposite direction (toward the pole). The induced north pole is on the side of the loop toward the magnet, and the two north poles repel one another, as in Fig. 35-5.

This effect occurs no matter what the sense of rotation of the original orbit, so the magnetization in a diamagnetic material always opposes the applied field. The ratio of the magnetization contribution to the field $\mu_0 M$ to the applied field B_0, given by $\kappa_m - 1$ according to Eq. 35-17, amounts to about -10^{-6} to -10^{-5} for typical diamagnetic materials. Table 35-3 shows some diamagnetic materials and their permeabilities.

Ferromagnetism

Ferromagnetism, like paramagnetism, occurs in materials in which the atoms have permanent magnetic dipole moments. What distinguishes ferromagnetic materials from paramagnetic materials is that in ferromagnetic materials there is a strong interaction between neighboring atoms that keeps their dipole moments aligned even when the external magnetic field is removed. Whether or not this occurs depends on the strength of the atomic dipoles and, because the dipole field changes with distance, on the separation between the atoms of the material. Certain atoms might be ferromagnetic in one kind of material but not in another, because their spacing is different. Familiar ferromagnetic materials at room temperature include the elements iron, cobalt, and nickel. Less familiar ferromagnetic elements, some of which show their ferromagnetism only at temperatures much below room temperature, are the elements of the rare earths, such as gadolinium or dysprosium. Compounds and alloys also may be ferromagnetic; for example, CrO_2, the basic ingredient of magnetic recording tape, is ferromagnetic even though neither of the elements chromium or oxygen is ferromagnetic at room temperature.

TABLE 35-3 **Relative Permeability of Some Diamagnetic Substances at Room Temperature**

Substance	$\kappa_m - 1$
Mercury	-3.2×10^{-5}
Silver	-2.6×10^{-5}
Bismuth	-1.7×10^{-5}
Ethyl alcohol	-1.3×10^{-5}
Copper	-9.7×10^{-6}
Carbon dioxide (1 atm)	-1.1×10^{-8}
Nitrogen (1 atm)	-5.4×10^{-9}

We can decrease the effectiveness of the coupling between neighboring atoms that causes ferromagnetism by increasing the temperature of a substance. The temperature at which a ferromagnetic material becomes paramagnetic is called its *Curie temperature.* The Curie temperature of iron, for instance, is 770°C; above this temperature, iron is paramagnetic. The Curie temperature of gadolinium metal is 16°C; at room temperature, gadolinium is paramagnetic, whereas at temperatures below 16°C, gadolinium becomes ferromagnetic.

The enhancement of the applied field in ferromagnets is considerable. The total magnetic field $\vec{\mathbf{B}}$ inside a ferromagnet may be 10^3 or 10^4 times the applied field $\vec{\mathbf{B}}_0$. The permeability κ_m of a ferromagnetic material is not a constant; neither the field $\vec{\mathbf{B}}$ nor the magnetization $\vec{\mathbf{M}}$ increases linearly with $\vec{\mathbf{B}}_0$, even at small values of $\vec{\mathbf{B}}_0$.

Hysteresis and Magnetic Domains. Let us insert a ferromagnetic material such as iron into a solenoid as in Fig. 35-8*b*. We assume that the current is initially zero and that the iron is unmagnetized, so that initially both B_0 and M are zero. We increase B_0 by increasing the current in the solenoid. The magnetization increases rapidly toward a saturation value as indicated in Fig. 35-11 by the segment *ab*. Now we decrease the current to zero. The magnetization does not retrace its original path, but instead the iron remains magnetized (at point *c*) even when the applied field B_0 is zero. If we then reverse the direction of the current in the solenoid, we reach a saturated magnetization in the opposite direction (point *d*), and returning the current to zero we find that the sample retains a permanent magnetization at point *e*. We can then increase the current again to return to the saturated magnetization in the original direction (point *b*). The path *bcdeb* can be repeatedly followed.

The behavior shown in Fig. 35-11 is called *hysteresis.* At points *c* and *e*, the iron is magnetized, even though there is no current in the solenoid. Furthermore, the iron "remembers" how it became magnetized, a negative current producing a magnetization different from a positive one. This "memory" is essential to the operation of magnetic storage of information, such as on cassette tapes or computer disks.

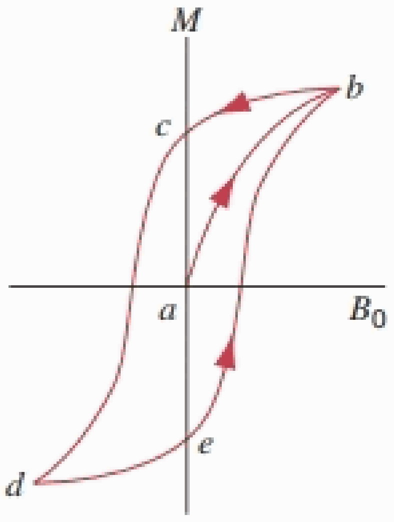


FIGURE 35-11. The variation of the magnetization of a sample of ferromagnetic material as the applied field is changed. The loop *bcdeb* is called a *hysteresis curve.*

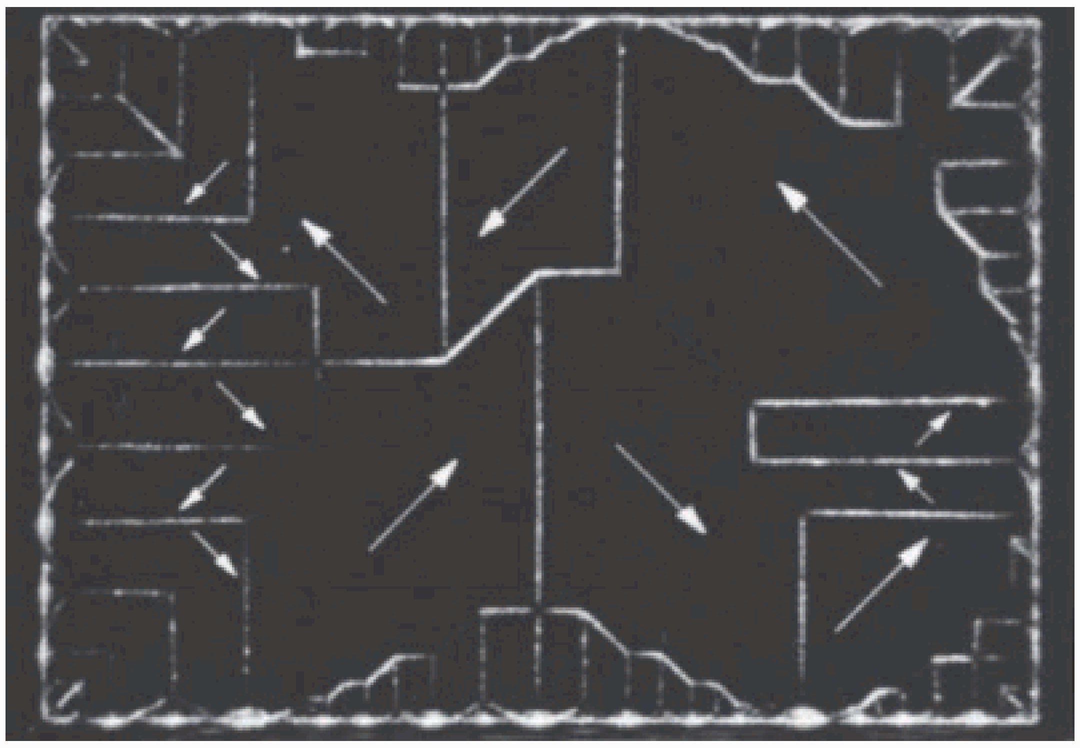


FIGURE 35-12. Domain patterns for a single crystal of nickel. The white lines, which show the boundaries of the domains, are produced by iron oxide powder sprinkled on the surface. The arrows illustrate the orientation of the magnetic dipoles within the domains.

The approach of a ferromagnet to saturation occurs through a mechanism different from that of a paramagnet (which we described by means of the rotation of individual magnetic dipoles into alignment with the applied field). A material such as iron is composed of a large number of microscopic crystals. Within each crystal are *magnetic domains,* regions of roughly 0.01 mm in size in which the coupling of atomic magnetic dipoles produces essentially perfect alignment of all the atoms. Figure 35-12 shows a pattern of domains in a single crystal of ferromagnetic nickel. There are many domains, each with its dipoles pointing in a different direction, and the net result of adding these dipole moments in an unmagnetized ferromagnet gives a magnetization of zero.

When the ferromagnet is placed in an external field, two effects may occur: (1) dipoles outside the walls of domains that are aligned with the field can rotate into alignment, in effect allowing such domains to grow at the expense of neighboring domains; and (2) the dipoles of nonaligned domains may swing entirely into alignment, with the applied field. In either case, there are now more dipoles aligned with the field, and the material has a large magnetization. When the field is removed, the domain walls do not move completely back to their former positions, and the material retains a magnetization in the direction of the applied field.

SAMPLE PROBLEM 35-4. A paramagnetic substance is composed of atoms with a magnetic dipole moment of 3.3 μ_B. It is placed in a magnetic field of strength 5.2 T. To what temperature must the substance be cooled so that the magnetic energy of each atom would be as large as the mean translational kinetic energy per atom?

Solution The magnetic energy of a dipole in an external field is $U = -\vec{\boldsymbol{\mu}} \cdot \vec{\mathbf{B}}$, and the mean translational kinetic energy per atom is $(3/2)kT$ (see Section 22-4). These are equal in magnitude when the temperature is

$$T = \frac{\mu B}{(3/2)k} = \frac{(3.3)(9.27 \times 10^{-24}\ \mathrm{J/T})(5.2\ \mathrm{T})}{(1.5)(1.38 \times 10^{-23}\ \mathrm{J/K})} = 7.7\ \mathrm{K}.$$

SAMPLE PROBLEM 35-5. Calculate the change in magnetic moment of a circulating electron in an applied field B_0 of 2.0 T acting perpendicular to the plane of the orbit. Take $r = 5.29 \times 10^{-11}$ m for the radius of the orbit, corresponding to the normal state of an atom of hydrogen.

Solution We can write Eq. 35-9 as

$$\mu = \tfrac{1}{2}erv = \tfrac{1}{2}er^2\omega,$$

using $v = r\omega$. The change $\Delta\mu$ in magnetic moment corresponding to a change in the angular frequency is then

$$\begin{aligned}\Delta\mu &= \tfrac{1}{2}er^2\,\Delta\omega = \tfrac{1}{2}er^2\left(\pm\frac{eB_0}{2m}\right) = \pm\frac{e^2B_0r^2}{4m}\\ &= \pm\frac{(1.6 \times 10^{-19}\ \mathrm{C})^2(2.0\ \mathrm{T})(5.29 \times 10^{-11}\ \mathrm{m})^2}{4(9.1 \times 10^{-31}\ \mathrm{kg})}\\ &= \pm 3.9 \times 10^{-29}\ \mathrm{J/T},\end{aligned}$$

where we have used Eq. 35-20 for $\Delta\omega$.

Compared with the value of the magnetic moment, $\mu_B = 9.27 \times 10^{-24}$ J/T, we see that this effect amounts to only about 4×10^{-6} of the magnetic moment. This is consistent with the order of magnitude expected for diamagnetic effects (Table 35-3).

35-6 THE MAGNETISM OF THE PLANETS (Optional)

Although magnetic compasses had already been in use as navigational instruments for several centuries, the explanation for their behavior was not well understood until 1600, when Sir William Gilbert, later physician to Queen Elizabeth I, proposed that the Earth is a huge magnet with a magnetic pole near each geographic pole. Subsequent researchers have carefully mapped the Earth's magnetic field, and interplanetary spacecraft have studied the magnetic fields of other planets.

The Earth's field can be considered roughly that of a magnetic dipole, with moment $\mu = 8.0 \times 10^{22}$ J/T. The field at the surface has a magnitude that ranges from about 30 μT near the equator to about 60 μT near the poles. For a dipole, we expect the magnetic field on the axis to be twice the field at the same distance along the bisector, as is the case for an electric dipole (see Section 26-3). These values for the Earth's magnetic field are consistent with these expectations.

The axis of the dipole makes an angle of about 11.5° with the Earth's rotational axis (which itself makes an angle of 23.5° with the normal to the plane of the Earth's orbit about the Sun, as shown in Fig. 35-13). What we

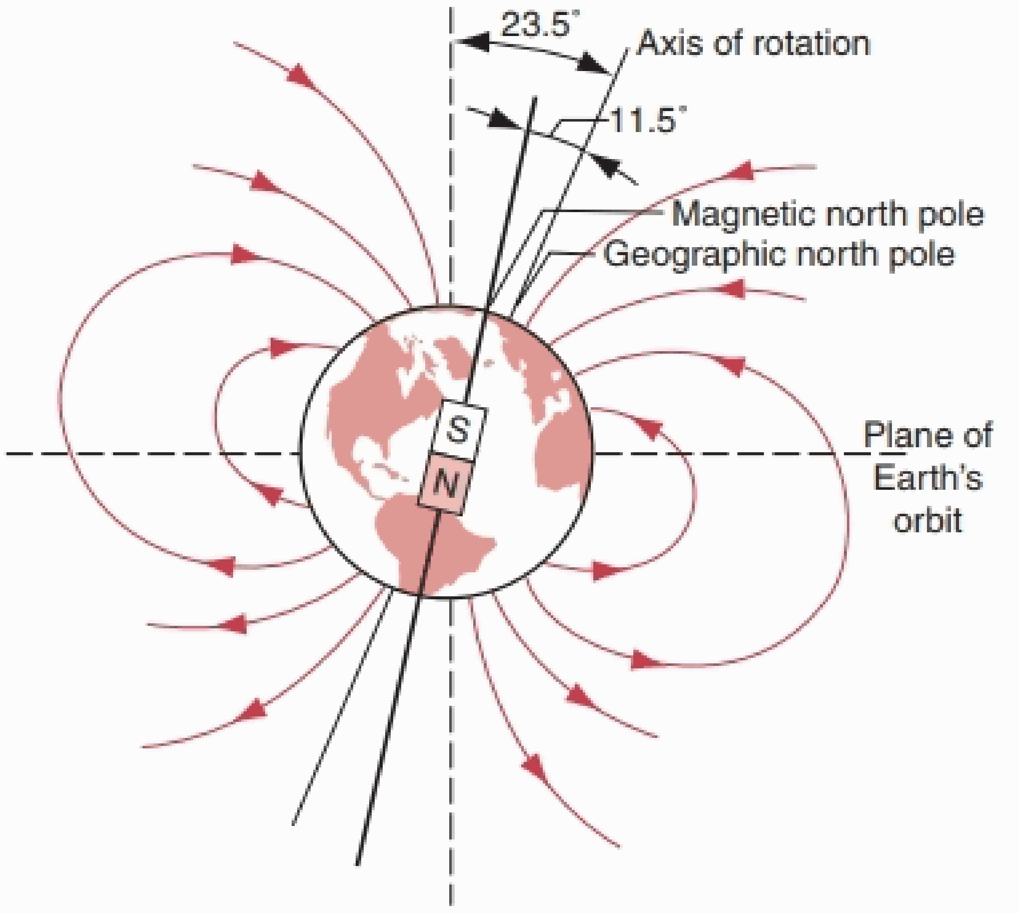


FIGURE 35-13. A simplified representation of the Earth's magnetic field near its surface. Note that the magnetic north pole is actually a south pole of the dipole that represents the Earth's field. The magnetic axis lies roughly halfway between the axis of rotation and the normal to the plane of the Earth's orbit (vertical dashed line).

FIGURE 35-14. The spectacular aurora borealis, also known as the "northern lights."

commonly call the north magnetic pole, which is located in northern Canada, is in fact the south pole of the Earth's dipole, as we have defined it by the converging of the magnetic field lines. The south magnetic pole, which is located in Antarctica, is represented by the north pole of a dipole, because the lines of $\vec{\mathbf{B}}$ emerge from it. Put another way, when you use a magnetic compass to tell direction, the end of the compass that points toward the north is a true north pole of the suspended magnet in your compass; it is attracted toward a true south pole, which is near the north geographical pole of Earth.

The Earth's magnetic field has practical importance not only in navigation but also in prospecting and in communications. It has therefore been studied extensively for many years, on the surface by measuring its magnitude and direction and above the surface by using orbiting satellites. Among its other effects are the Van Allen radiation belts that surround Earth (see Fig. 32-15) and the so-called "northern lights," the brilliant display of the aurora* (Fig. 35-14).

Because we find magnetized rocks in the ground, it is tempting to suggest a core of permanently magnetized rocks as the source of the Earth's magnetic field. However, this cannot be correct, because the temperature of the core is several thousand degrees, far above the Curie temperature of iron. Iron in the Earth's core therefore cannot be ferromagnetic.

Furthermore, from measurements over the past few hundred years we know that the north magnetic pole migrates relative to the north geographic pole, and from the geologic record we know that the poles reverse on a time scale of several hundred thousand years. (Moreover, as we discuss later, some planets in the solar system that have compositions similar to Earth's have no magnetic field, whereas other planets that certainly contain no magnetic material have very large fields.) Such observations are difficult to explain based on the assumption of a permanently magnetized core.

The exact source of the Earth's magnetism is not completely understood, but it probably involves some sort of *dynamo* effect. The outer core contains minerals in a liquid state, which easily conduct electricity. A small initial magnetic field causes currents to flow in this moving conductor, by Faraday's law of induction. These currents may enhance the magnetic field, and this enhanced field is what we observe as the Earth's field. However, we know from our study of induction that a conductor moving in a magnetic field experiences a braking force. The source of the energy needed to overcome the braking force and keep the core moving is not yet understood.

The Earth contains a record of changes in both the direction and the magnitude of the field. Ancient pottery samples, for example, contain tiny iron particles, which became magnetized in the Earth's field as the pottery was cooled after its firing. From the strength of the magnetization of the particles, we can deduce the intensity of the Earth's field at the time and place of the firing. A geological record of similar origin is preserved in the ocean floor (Fig. 35-15). As molten magma oozes from a ridge and solidifies, the iron particles become magnetized. The direction of magnetization of the particles shows the direction of the

---

*See "The Dynamic Aurora," by Syun-Ichi Akasofu, *Scientific American,* May 1989, p. 90.

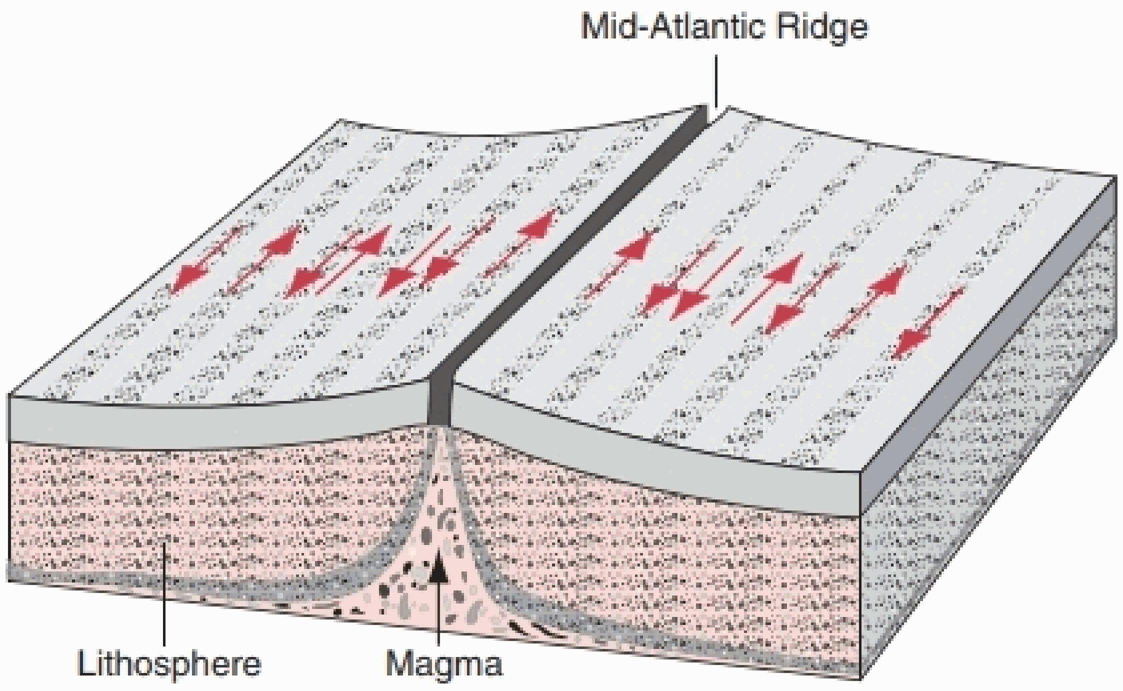


FIGURE 35-15. As molten material emerges through a ridge in the ocean floor and cools, it preserves a record of the direction of the Earth's magnetic field at that time (arrows). Each segment might represent a time of 100,000 to 1,000,000 years.

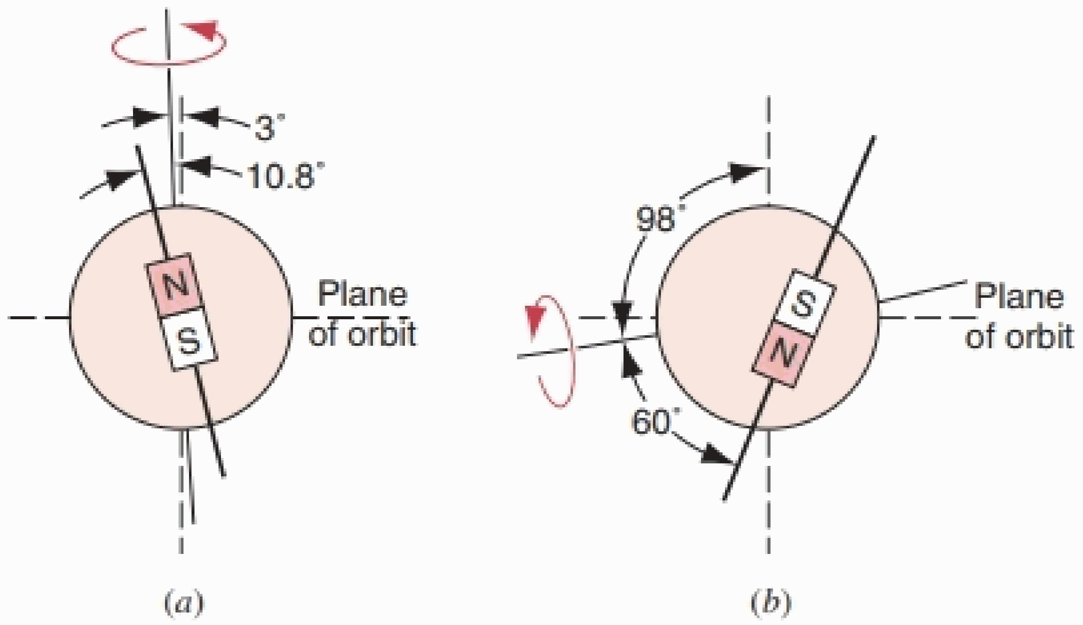


FIGURE 35-16. (*a*) The alignment of the magnetic dipole axis of Jupiter relative to its axis of rotation and the plane of its orbit. Note that, in contrast to the Earth, the north magnetic pole is a true north pole of the dipole field. (*b*) The alignment of the magnetic dipole axis of Uranus.

Earth's field. From the patterns of magnetization, we can deduce that the Earth's poles have reversed fairly regularly over geologic history. This reversal occurs about every 100,000–1,000,000 years and has become more frequent in recent times. The reasons for these reversals and their accelerating rate are not known but presumably involve the dynamo effect in some way.*

In recent years, interplanetary space probes have been able to measure the direction and magnitude of the magnetic fields of the planets. These observations support the dynamo mechanism as the source of these fields. Table 35-4 shows values of the magnetic dipole moments and surface magnetic fields of the planets.

Venus, whose core is similar to Earth's, has no field because its rotation is too slow (once every 244 Earth days) to sustain the dynamo effect. Mars, whose rotational period is nearly the same as Earth's, has only a tiny field because its core is presumably very small, a fact deduced from the measured mean density of Mars. The outer planets (Jupiter and beyond) are composed mostly of hydrogen and helium, which ordinarily are not expected to be magnetic; however, at the high pressures and temperatures near the center of these planets, hydrogen and helium can behave like metals, in particular showing large electrical conductivity and permitting the dynamo effect.

Figure 35-16 shows the alignment of the rotational axis and magnetic field axis of Jupiter and Uranus; compare these with the Earth shown in Fig. 35-13. Note that the rotational axis of Uranus is nearly parallel to the plane of its orbit, in contrast to the other planets. Notice also that the magnetic axis of Uranus is badly misaligned with its rotational axis and that the dipole is displaced from the center of the planet. A similar situation occurs for the planet Neptune. Unfortunately, our observational information on the planets is limited to that gathered from space flights that were in the neighborhood of the planet only for a day or so. If we could examine their other physical properties and their geologic records, we would learn a great deal more about the origin of planetary magnetism.*

TABLE 35-4 **Magnetic Fields in the Solar System**

Planet	μ (A · m^2)	*B at Surface* (μT)
Mercury	5×10^{19}	0.35
Venus	$< 10^{19}$	< 0.01
Earth	8.0×10^{22}	30
Mars	1.5×10^{19}	0.04
Jupiter	1.6×10^{27}	430
Saturn	4.7×10^{25}	20
Uranus	4.0×10^{24}	30
Neptune	2.2×10^{24}	20

SAMPLE PROBLEM 35-6. A measurement of the horizontal component B_h of the Earth's field at the location of Tucson, Arizona, gave a value of 26 μT. By suspending a small magnet like a compass that is free to swing in a vertical plane, it is possible to measure the angle between the field direction and the horizontal plane, called the *inclination* or the *dip angle* ϕ_i. The dip

---

*See "The Evolution of the Earth's Magnetic Field," by Jeremy Bloxham and David Gubbins, *Scientific American,* December 1989, p. 68; and "The Source of the Earth's Magnetic Field," by Charles R. Carrigan and David Gubbins, *Scientific American,* February 1979, p. 118.

*See "Magnetic Fields in the Cosmos," by E. N. Parker, *Scientific American,* August 1983, p. 44; and "Uranus," by Andrew P. Ingersoll, *Scientific American,* January 1987, p. 38.

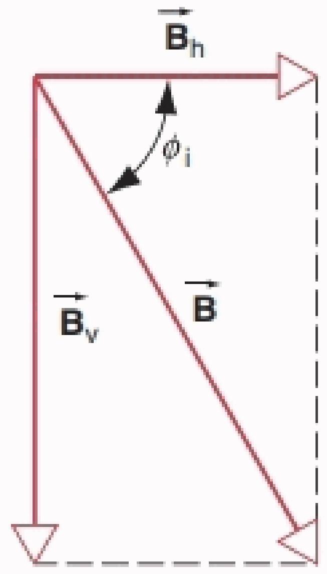


FIGURE 35-17. Sample Problem 35-6. The horizontal and vertical components of the Earth's magnetic field near Tucson, Arizona. The angle ϕ_i is the dip angle.

angle at Tucson was measured to be 59°. Find the magnitude of the field and its vertical component at that location.

Solution As Fig. 35-17 shows, the magnitude of the field can be found from

$$B = \frac{B_h}{\cos \phi_i} = \frac{26\ \mu\text{T}}{\cos 59^\circ} = 50\ \mu\text{T}.$$

The vertical component is given by

$$B_v = B_h \tan \phi_i = (26\ \mu\text{T})(\tan 59^\circ) = 43\ \mu\text{T}.$$

As expected for a dipole field (see Fig. 35-13), measured values of the dip angle range from 0° near the equator (actually, the *magnetic* equator) to 90° near the magnetic poles.

35-7 GAUSS' LAW FOR MAGNETISM

Figure 35-18*a* shows the electric field associated with an insulating rod having equal quantities of positive and negative charge placed on opposite ends. This is an example of an electric dipole. Figure 35-18*b* shows the analogous case of the magnetic dipole, such as the familiar bar magnet, with a north pole at one end and a south pole at the other end. At this level the electric and magnetic cases look quite similar. (Compare Fig. 26-14*b* with Fig. 32-6 to see another illustration of this similarity.) In fact, we might be led to postulate the existence of individual magnetic poles analogous to electric charges; such poles, if they existed, would produce magnetic fields (similar to electric fields produced by charges) proportional to the strength of the pole and inversely proportional to the square of the distance from the pole. As we shall see, this hypothesis is not consistent with experiment.

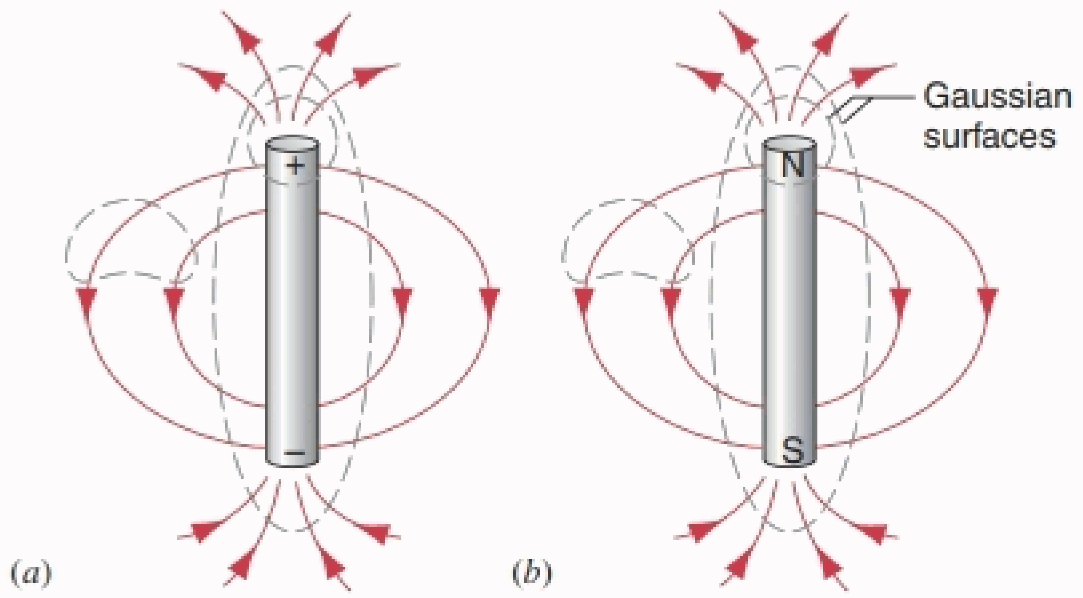


FIGURE 35-18. (*a*) An electric dipole, consisting of an insulating rod with a positive charge at one end and a negative charge at the other. Several Gaussian surfaces are shown. (*b*) A magnetic dipole, consisting of a bar magnet with a north pole at one end and a south pole at the other.

Let us cut the objects of Fig. 35-18 in half and separate the two pieces. Figure 35-19 shows that the electric and magnetic cases are no longer similar. In the electric case, we have two objects that, if separated by a sufficiently large distance, could be regarded as point charges of opposite polarities, each producing a field characteristic of a point charge. In the magnetic case, however, we obtain not isolated north and south poles but instead a pair of magnets, each with its own north and south poles.

This appears to be an important difference between electric and magnetic dipoles: an electric dipole can be separated into its constituent single charges (or "poles"), but a magnetic dipole cannot. Each time we try to divide a magnetic dipole into separate north and south poles, we create a new pair of poles. This process occurs microscopically, down to the level of individual atoms. Each atom behaves like a magnetic dipole having a north and a south pole, and as far as we yet know the dipole, rather than the single isolated pole, appears to be the smallest fundamental unit of magnetic structure.

This difference between electric and magnetic fields has a mathematical expression in the form of Gauss' law. In Fig. 35-18*a*, the flux of the electric field through the different Gaussian surfaces depends on the net charge enclosed

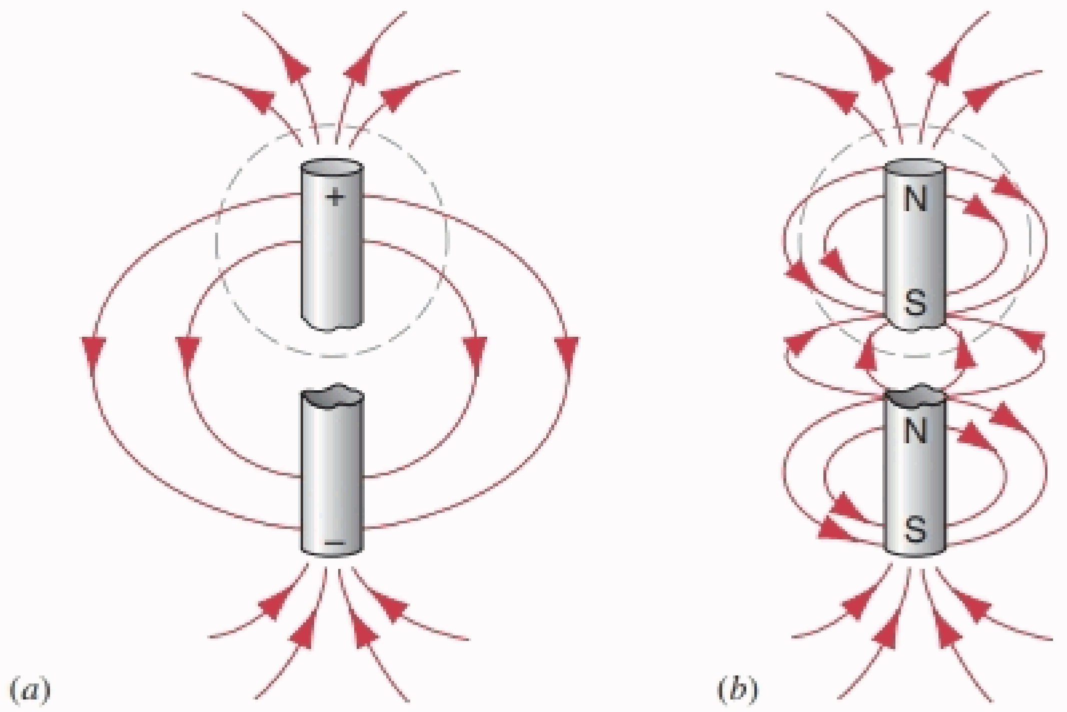


FIGURE 35-19. (*a*) When the electric dipole of Fig. 37-18*a* is cut in half, the positive charge is isolated on one piece and the negative charge on the other. (*b*) When the magnetic dipole of Fig. 35-18*b* is cut in half, a new pair of north and south poles appears. Note the difference in the field patterns.

by each surface. If the surface encloses no charge at all, or no net charge (that is, equal quantities of positive and negative charge, such as the entire dipole), the flux of the electric field vector through the surface is zero. If the surface cuts through the dipole, so that it encloses a net charge q, the flux Φ_E of the electric field is given by Gauss' law:

$$\Phi_E = \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q/\epsilon_0. \qquad (35\text{-}21)$$

We can similarly construct Gaussian surfaces for the magnetic field, as in Fig. 35-18*b*. If the Gaussian surface contains no net "magnetic charge," the flux Φ_B of the magnetic field through the surface is zero. However, as we have seen, even those Gaussian surfaces that cut through the bar magnet enclose no net magnetic charge, because every cut through the magnet gives a piece having both a north and a south pole. The magnetic form of Gauss' law is written

$$\Phi_B = \oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0. \qquad (35\text{-}22)$$

The net flux of the magnetic field through any closed surface is zero.

Figure 35-20 shows a more detailed representation of the magnetic fields of a bar magnet and a solenoid, both of which can be considered as magnetic dipoles. Note in Fig. 35-20*a* that lines of $\vec{\mathbf{B}}$ enter the Gaussian surface inside the magnet and leave it outside the magnet. The total inward flux equals the total outward flux, and the net flux Φ_B for the surface is zero. The same is true for the Gaussian surface through the solenoid shown in Fig. 35-20*b*. In neither case is there a single point from which the lines of $\vec{\mathbf{B}}$ originate or to which they converge; that is, *there is no isolated magnetic charge.*

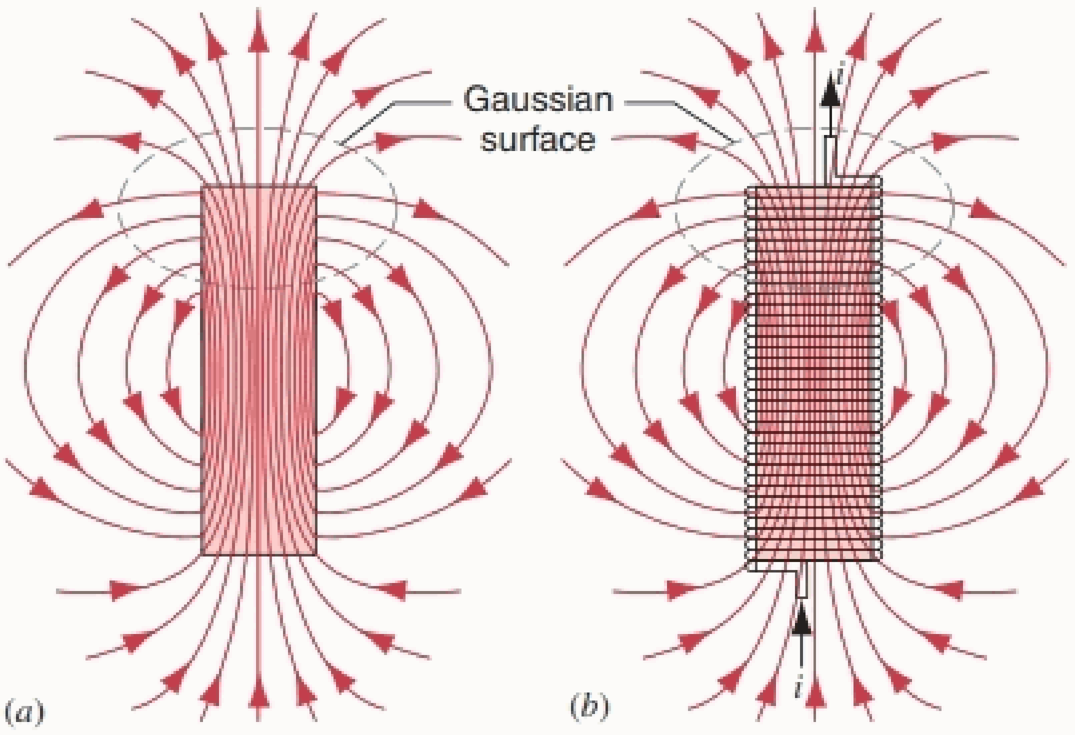


FIGURE 35-20. Lines of $\vec{\mathbf{B}}$ for (*a*) a bar magnet and (*b*) a short solenoid. In each case, the north pole is at the top of the figure. The dashed lines represent Gaussian surfaces.

Magnetic Monopoles

We showed in Chapter 27 that Gauss' law for electric fields is equivalent to Coulomb's law, which is based on the experimental observation of the force between point charges. Gauss' law for magnetism is also based on an experimental observation: the failure to observe isolated magnetic poles, such as a single north pole or south pole.

The existence of isolated magnetic charges was proposed in 1931 by theoretical physicist Paul Dirac on the basis of arguments using quantum mechanics and symmetry. He called those charges *magnetic monopoles* and derived some basic properties expected of them, including the magnitude of the "magnetic charge" (analogous to the electronic charge e). Following Dirac's prediction, searches for magnetic monopoles were made using large particle accelerators as well as by examining samples of terrestrial and extraterrestrial matter. None of these early searches turned up any evidence for the existence of magnetic monopoles.

Recent attempts to unify the laws of physics, bringing together the strong, weak, and electromagnetic forces into a single framework, have reawakened interest in magnetic monopoles. These theories predict the existence of extremely massive magnetic monopoles, roughly 10^{16} times the mass of the proton. This is certainly far too massive to be made in any accelerator on Earth; in fact, the only known conditions under which such monopoles could have been made would have occurred in the hot, dense matter of the early universe. Searches for magnetic monopoles continue to be made, but convincing evidence for their existence has not yet been obtained.* For the present, we assume either that monopoles do not exist, so that Eq. 35-22 is exactly and universally valid, or else that if they do exist they are so exceedingly rare that Eq. 35-22 is a highly accurate approximation. Equation 35-22 then assumes a fundamental role as a description of the behavior of magnetic fields in nature, and it is included as one of the four Maxwell equations of electromagnetism.

---

*See "Searches for Magnetic Monopoles and Fractional Electric Charges," by Susan B. Felch, *The Physics Teacher,* March 1984, p. 142. See also "Superheavy Magnetic Monopoles," by Richard A. Carrigan, Jr. and W. Peter Trower, *Scientific American,* April 1982, p. 106.

MULTIPLE CHOICE

35-1 The Magnetic Dipole

1. A magnetic dipole is oriented in a uniform magnetic field so that the potential energy is a maximum. The magnitude of the torque on this dipole will be

(A) a maximum. (B) a minimum.
(C) dependent on the source of the magnetic dipole.

2. A simple bar magnet hangs from a string as in Fig. 35-21*a*. A uniform magnetic field $\vec{B}$ directed horizontally to the right is then established. Which of the pictures in Fig. 35-21*b* shows the equilibrium orientation of the bar magnet?

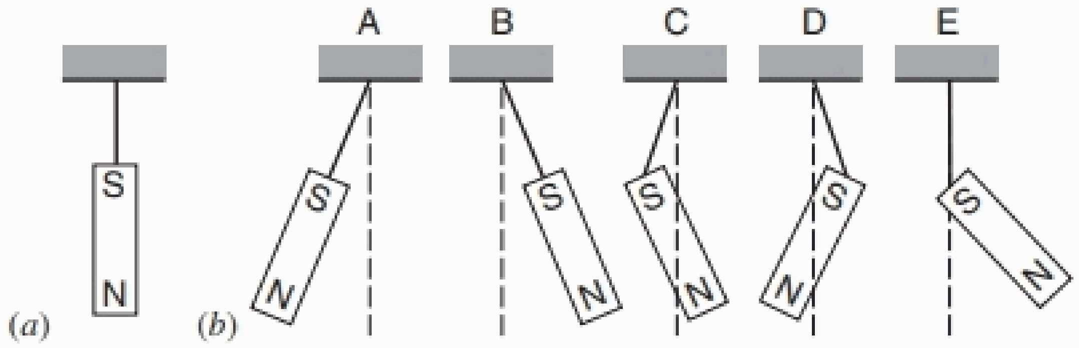


FIGURE 35-21. Multiple-choice question 2.

35-2 The Force on a Dipole in a Nonuniform Field

3. A magnetic dipole is in a nonuniform magnetic field. The dipole will

(A) always be attracted to the region with a stronger magnetic field.
(B) always be repelled from the region with a stronger magnetic field.
(C) might be repelled, but will eventually flip over and be attracted to the region with a stronger magnetic field.
(D) might be attracted, but could flip over and then be repelled from the region with a stronger magnetic field.

4. For distances d much larger than the dimensions of the magnetic dipoles, the force between two dipoles is proportional to

(A) d^{-6}. (B) d^{-4}. (C) d^{-3}. (D) d^{-2}.

35-3 Atomic and Nuclear Magnetism

5. Which of the following does not significantly contribute to the magnetic properties of a substance?

(A) Orbital magnetic moments of electrons
(B) Spin magnetic moments of electrons
(C) Magnetic moments of protons and neutrons
(D) All contribute equally.

35-4 Magnetization

6. What are units for the magnetization $\vec{M}$?

(A) T (B) T/m^3 (C) C/m·s (D) C·m/s

35-5 Magnetic Materials

7. Which type of substance has the largest atomic dipole moments?

(A) Paramagnetic (B) Diamagnetic
(C) Ferromagnetic
(D) Paramagnetic and ferromagnetic substances tend to be about the same, whereas diamagnetic substances have no permanent magnetic moment.
(E) All three types are about the same.

8. What happens when a paramagnetic substance is placed in an external magnetic field?

(A) The atomic dipole moments weaken slightly, but tend to line up with the external field.
(B) The atomic dipole moments strengthen slightly, and tend to line up with the external field.
(C) The atomic dipole moments weaken slightly, and tend to line up against the external field.
(D) The atomic dipole moments strengthen slightly, but tend to line up against the external field.

9. A small, cylindrical bar magnet has a diameter of 1 cm, a length of 2 cm, and a dipole moment of 5 J/T. Assuming the magnetic field were produced by a single current loop around the magnet, about how large would this current be?

(A) 0.7 mA (B) 0.7 A (C) 70 A (D) 70,000 A

35-6 The Magnetism of the Planets

10. Assume that the Earth's magnetic dipole moment is caused by a ring of charge on the equator and the Earth's rotation on its axis.

(*a*) What would be the sign of this charge?

(A) Positive (B) Negative
(C) Either answer would produce the same dipole moment.

(*b*) Which estimate is closest to the amount of charge required?

(A) 10^{15} C (B) 10^{5} C (C) 10^{-5} C (D) 10^{-15} C

35-7 Gauss' Law for Magnetism

11. If a magnetic monopole were to exist, what would be appropriate units?

(A) Wb/T (B) T/m^3 (C) C·m/s (D) C/T
(E) The charge of the monopole would require the introduction of a new unit.

12. A (hypothetical?) magnetic monopole passes through a wire loop. Which of the graphs in Fig. 35-22 shows the current as a function of time through the loop?

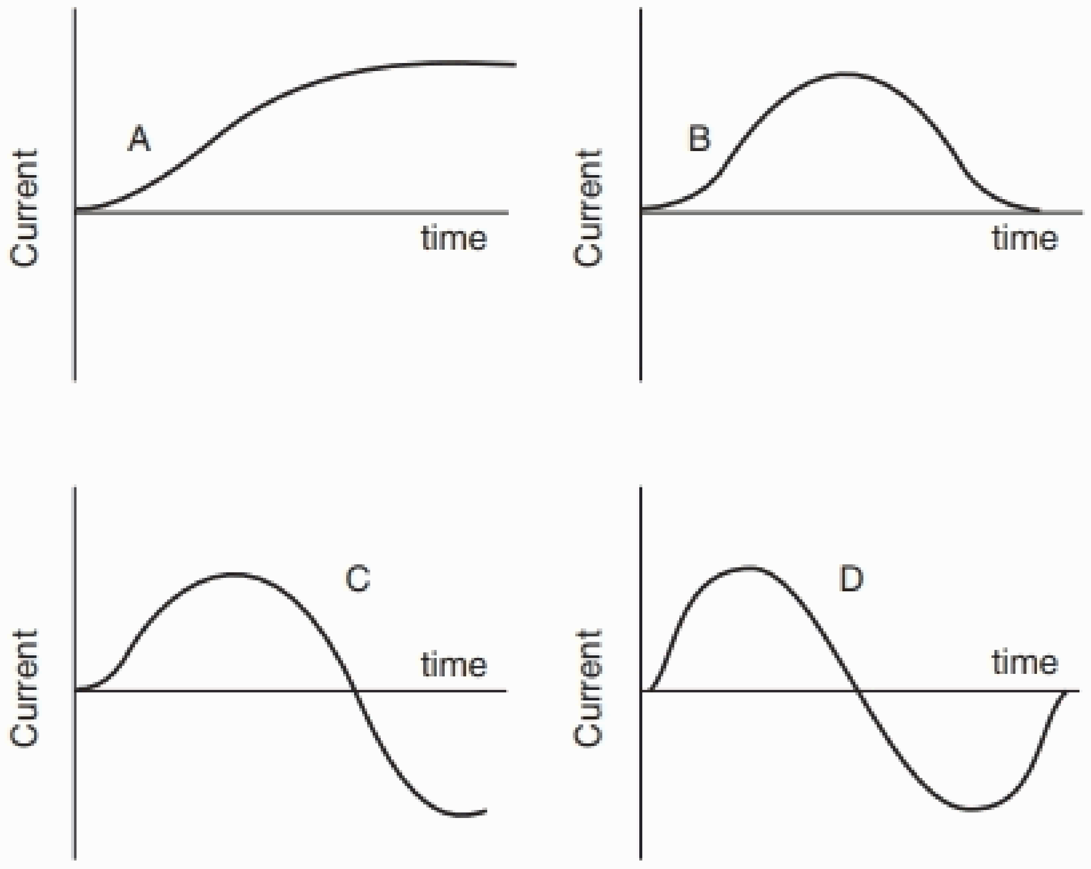


FIGURE 35-22. Multiple-choice question 12.

QUESTIONS

1. In Sample Problem 35-1 we showed that the work required to turn a current loop end for end in an external magnetic field is $2\mu B$. Does this result hold no matter what the original orientation of the loop was?
2. The torque exerted by a magnetic field on a magnetic dipole can be used to measure the strength of that magnetic field. For an accurate measurement, does it matter whether the dipole moment is small or not? Recall that, in the case of measurement of an electric field, the test charge was to be as small as possible so as not to disturb the source of the field.
3. You are given a frictionless sphere the size of a Ping-Pong ball and told that it contains a magnetic dipole. What experiments would you carry out to find the magnitude and the direction of its magnetic dipole moment?
4. How might you measure the magnetic dipole moment of a compass needle?
5. A circular loop of wire lies on the floor of the room in which you are sitting. It carries a constant current i in a clockwise sense, viewed from above. What is the direction of the magnetic dipole moment of this current loop?
6. Two iron bars are identical in appearance. One is a magnet and one is not. How can you tell them apart? You are not permitted to suspend either bar as a compass needle or to use any apparatus other than the two bars.
7. Two iron bars always attract, regardless of the combination in which their ends are brought near each other. Can you conclude that one of the bars must be unmagnetized?
8. How are these phenomena similar and different? (*a*) A charged rod can attract small pieces of uncharged insulators. (*b*) A permanent magnet can attract any nonmagnetized sample of ferromagnetic material.
9. How can you determine the polarity of an unlabeled magnet?
10. Show that, classically, a spinning positive charge will have a spin magnetic moment that points in the same direction as its spin angular momentum.
11. The neutron, which has no charge, has a magnetic dipole moment. Is this possible on the basis of classical electromagnetism, or does this evidence alone indicate that classical electromagnetism has broken down?
12. Must all permanent magnets have identifiable north and south poles? Consider geometries other than the bar or horseshoe magnet.
13. Consider these two situations: (*a*) a (hypothetical) magnetic monopole is pulled through a single-turn conducting loop along its axis, at a constant speed; (*b*) a short bar magnet (a magnetic dipole) is similarly pulled. Compare qualitatively the net amounts of charge transferred through any cross section of the loop during these two processes. Experiments designed to detect possible magnetic monopoles exploit such differences.
14. A certain short iron rod is found, by test, to have a north pole at each end. You sprinkle iron filings over the rod. Where (in the simplest case) will they cling? Make a rough sketch of what the lines of $\vec{\mathbf{B}}$ must look like, both inside and outside the rod.
15. Starting with $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ in the positions and orientations shown in Fig. 35-23, with $\vec{\mathbf{X}}$ fixed but $\vec{\mathbf{Y}}$ free to rotate, what happens if (*a*) $\vec{\mathbf{X}}$ is an electric dipole and $\vec{\mathbf{Y}}$ is a magnetic dipole; (*b*) $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ are both magnetic dipoles; (*c*) $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ are both electric dipoles? Answer the same questions if $\vec{\mathbf{Y}}$ is fixed and $\vec{\mathbf{X}}$ is free to rotate.

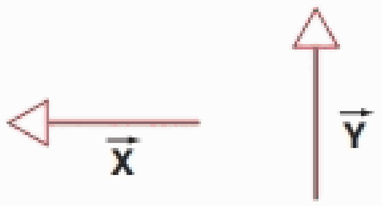


FIGURE 35-23. Question 15.

16. You are a manufacturer of compasses. (*a*) Describe ways in which you might magnetize the needles. (*b*) The end of the needle that points north is usually painted a characteristic color. Without suspending the needle in the Earth's field, how might you find out which end of the needle to paint? (*c*) Is the painted end a north or a south magnetic pole?
17. Would you expect the magnetization at saturation for a paramagnetic substance to be very much different from that for a saturated ferromagnetic substance of about the same size? Why or why not?
18. Can you give a reason for the fact that ferromagnetic materials become purely paramagnetic at depths greater than about 20 km below the Earth's surface?
19. It is desired to demagnetize a sample of ferromagnetic material that retains the magnetism acquired when placed in an external field. Must the temperature of the sample be raised to the melting temperature to accomplish this?
20. The magnetization induced in a given diamagnetic sphere by a given external magnetic field does not vary with temperature, in sharp contrast to the situation in paramagnetism. Explain this behavior in terms of the description that we have given of the origin of diamagnetism.
21. Explain why a magnet attracts an unmagnetized iron object such as a nail.
22. Does any net force or torque act on (*a*) an unmagnetized iron bar or (*b*) a permanent bar magnet when placed in a uniform magnetic field?
23. A nail is placed at rest on a frictionless tabletop near a strong magnet. It is released and attracted to the magnet. What is the source of the kinetic energy that it has just before it strikes the magnet?
24. Superconductors are said to be perfectly diamagnetic. Explain.
25. Explain why a small bar magnet that is placed vertically above a bowl made of superconducting lead needs no contact forces to support it.
26. Compare the magnetization curves for a paramagnetic substance (see Fig. 35-10) and for a ferromagnetic substance (see Fig. 35-11). What would a similar curve for a diamagnetic substance look like?

27. Why do iron filings line up with a magnetic field? After all, they are not intrinsically magnetized.

28. The Earth's magnetic field can be represented closely by that of a magnetic dipole located at or near the center of the Earth. The Earth's magnetic poles can be thought of as (*a*) the points where the axis of this dipole passes through the Earth's surface or (*b*) the points on the Earth's surface where a dip needle would point vertically. Are these necessarily the same points?

29. Are there any locations in the United States where a magnetic compass will point to the geographic north pole?

30. A "friend" borrows your favorite compass and paints the entire needle red. When you discover this you are lost in a cave and have with you two flashlights, a few meters of wire, and (of course) this book. How might you discover which end of your compass needle is the north-seeking end?

31. How can you magnetize an iron bar if the Earth is the only magnet around?

32. How would you go about shielding a certain volume of space from constant external magnetic fields? If you think it cannot be done, explain why.

33. Cosmic rays are charged particles that strike our atmosphere from some external source. We find that more low-energy cosmic rays reach the Earth near the north and south magnetic poles than at the (magnetic) equator. Why is this so?

34. Diamagnetic substances are repelled by magnetic fields. Why does a diamagnetic substance not "flip-over" and get attracted like any other magnetic dipole?

35. How might the magnetic dipole moment of the Earth be measured?

36. Give three reasons for believing that the flux Φ_B of the Earth's magnetic field is greater through the boundaries of Alaska than through those of Texas.

37. Aurorae are most frequently observed, not at the north and south magnetic poles, but at magnetic latitudes about 23° away from these poles (passing through Hudson Bay, for example, in the northern geomagnetic hemisphere). Can you think of any reason, however qualitative, why the auroral activity should not be strongest at the poles themselves?

38. Can you think of a mechanism by which a magnetic storm—that is, a strong perturbation of the Earth's magnetic field—can interfere with radio communication?

EXERCISES

35-1 The Magnetic Dipole

1. The magnetic dipole moment of the Earth is 8.0×10^{22} J/T. Assume that this is produced by charges flowing in the molten outer core of the Earth. If the radius of the circular path is 3500 km, calculate the required current.

2. A circular coil of 160 turns has a radius of 1.93 cm. (*a*) Calculate the current that results in a magnetic moment of 2.33 $\text{A}\cdot\text{m}^2$. (*b*) Find the maximum torque that the coil, carrying this current, can experience in a uniform 34.6-mT magnetic field.

3. Two concentric circular loops, radii 20.0 and 30.0 cm, in the *xy* plane each carry a clockwise current of 7.00 A, as shown in Fig. 35-24. (*a*) Find the net magnetic moment of this system. (*b*) Repeat if the current in the outer loop is reversed.

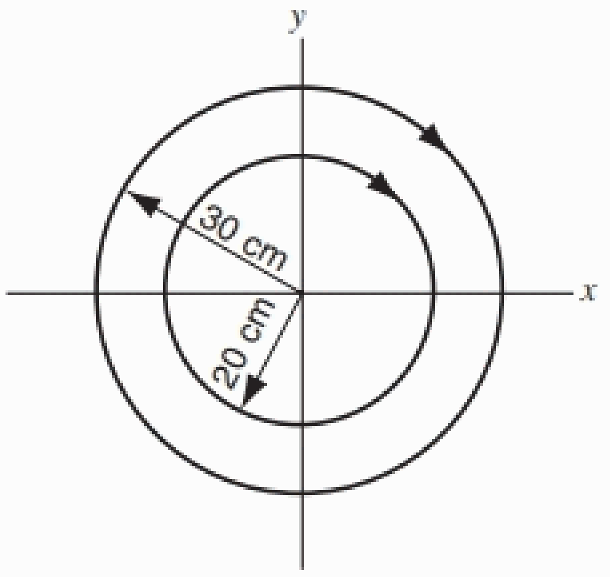


FIGURE 35-24. Exercise 3.

4. A circular wire loop whose radius is 16.0 cm carries a current of 2.58 A. It is placed so that the normal to its plane makes an angle of 41.0° with a uniform magnetic field of 1.20 T. (*a*) Calculate the magnetic dipole moment of the loop. (*b*) Find the torque on the loop.

5. The magnetic field B for various points on the axis of a square current loop of side a is given Problem 4 of Chapter 33. (*a*) Show that the axial field for this loop for $z \gg a$ is that of a magnetic dipole. (*b*) Find the magnetic dipole moment of this loop.

6. A circular loop of wire having a radius of 8.0 cm carries a current of 0.20 A. A unit vector parallel to the dipole moment $\vec{\mu}$ of the loop is given by $0.60\hat{\mathbf{i}} - 0.80\hat{\mathbf{j}}$. If the loop is located in a magnetic field given by $\vec{\mathbf{B}} = (0.25\hat{\mathbf{i}} + 0.30\hat{\mathbf{k}})$ T, find (*a*) the torque on the loop and (*b*) the magnetic potential energy of the loop.

7. You are given a closed circuit with radii a and b, as shown in Fig. 35-25, carrying current i. Find the magnetic dipole moment of the circuit.

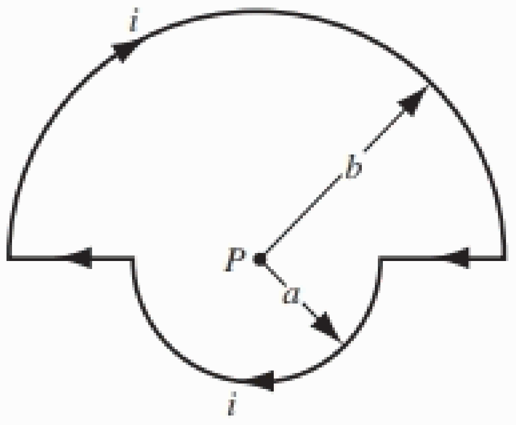


FIGURE 35-25. Exercise 7.

8. Consider the rectangular loop carrying current i shown in Fig. 35-26. Point P is located a distance x from the center of the loop. Find an expression for the magnetic field at P due to the current loop, assuming that P is very far away. With $\mu = iA = iab$, obtain an expression similar to Eq. 35-4 for the field due to a distant dipole at points in the plane of the loop (perpendicular to the axis). (Hint: Opposite sides of the rectangle can be treated together, but consider carefully the directions of $\vec{\mathbf{B}}$ due to each side.)

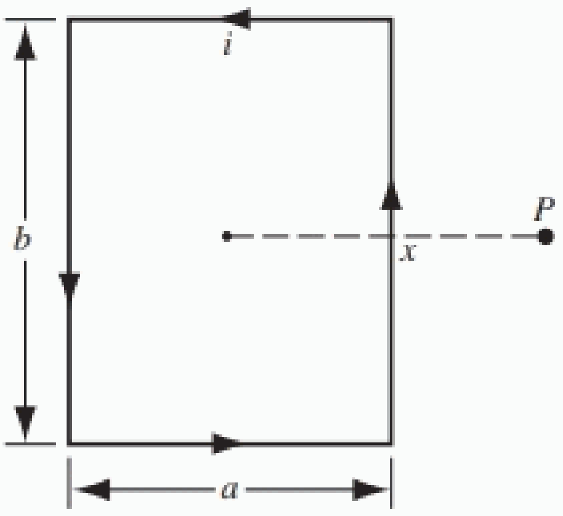


FIGURE 35-26. Exercise 8.

35-2 The Force on a Dipole in a Nonuniform Field

35-3 Atomic and Nuclear Magnetism

9. In the lowest energy state of the hydrogen atom, the most probable distance between the single orbiting electron and the central proton is 5.29×10^{-11} m. Calculate (*a*) the electric field and (*b*) the magnetic field set up by the proton at this distance, measured along the proton's axis of spin. See Table 35-1 for the magnetic moment of the proton.

10. Suppose that the hydrogen nuclei (protons) in 1.50 g of water could all be aligned. Calculate the magnetic field that would be produced 5.33 m from the sample along the alignment axis.

11. A charge q is distributed uniformly around a thin ring of radius r. The ring is rotating about an axis through its center and perpendicular to its plane at an angular speed ω. (*a*) Show that the magnetic moment due to the rotating charge is

$$\mu = \tfrac{1}{2} q\omega r^2.$$

(*b*) If L is the angular momentum of the ring, show that

$$\mu/L = q/2m.$$

35-4 Magnetization

12. The dipole moment associated with an atom of iron in an iron bar is $2.22\ \mu_B$. Assume that all the atoms in the bar, which is 4.86 cm long and has a cross-sectional area of 1.31 cm^2, have their dipole moments aligned. (*a*) What is the dipole moment of the bar? (*b*) What torque must be exerted to hold this magnet at right angles to an external field of 1.53 T?

13. A magnet in the shape of a cylindrical rod has a length of 4.8 cm and a diameter of 1.1 cm. It has a uniform magnetization of 5.3 kA/m. Calculate its magnetic dipole moment.

14. A solenoid with 16 turns/cm carries a current of 1.3 A. (*a*) By how much does the magnetic field inside the solenoid increase when a close-fitting chromium rod is inserted? (*b*) Find the magnetization of the rod. (See Table 35-2.)

35-5 Magnetic Materials

15. A 0.50-T magnetic field is applied to a paramagnetic gas whose atoms have an intrinsic magnetic dipole moment of 1.2×10^{-23} J/T. At what temperature will the mean kinetic energy of translation of the gas atoms be equal to the energy required to reverse such a dipole end for end in this magnetic field?

16. Measurements in mines and boreholes indicate that the temperature in the Earth increases with depth at the average rate of 30 C°/km. Assuming a surface temperature of 20°C, at what depth does iron cease to be ferromagnetic? (The Curie temperature of iron varies very little with pressure.)

17. A sample of the paramagnetic salt to which the magnetization curve of Fig. 35-10 applies is held at room temperature (300 K). At what applied magnetic field would the degree of magnetic saturation of the sample be (*a*) 50%? (*b*) 90%? (*c*) Are these fields attainable in the laboratory?

18. A sample of the paramagnetic salt to which the magnetization curve of Fig. 35-10 applies is immersed in a magnetic field of 1.8 T. At what temperature would the degree of magnetic saturation of the sample be (*a*) 50% and (*b*) 90%?

19. The paramagnetic salt to which the magnetization curve of Fig. 35-10 applies is to be tested to see whether it obeys Curie's law. The sample is placed in a 0.50-T magnetic field that remains constant throughout the experiment. The magnetization M is then measured at temperatures ranging from 10 to 300 K. Would Curie's law be found valid under these conditions?

20. The saturation magnetization of the ferromagnetic metal nickel is 511 kA/m. Calculate the magnetic moment of a single nickel atom. (Obtain needed data from Appendix D.)

21. The coupling mentioned in Section 35-5 as being responsible for ferromagnetism is not the mutual magnetic interaction energy between two elementary magnetic dipoles. To show this, calculate (*a*) the magnetic field a distance of 10 nm away along the dipole axis from an atom with magnetic dipole moment 1.5×10^{-23} J/T (cobalt) and (*b*) the minimum energy required to turn a second identical dipole end for end in this field. Compare with the results of Sample Problem 35-4. What do you conclude?

35-6 The Magnetism of the Planets

22. In Sample Problem 35-6 the vertical component of the Earth's magnetic field in Tucson, Arizona, was found to be 43 μT. Assume that this is the average value for all of Arizona, which has an area of 295,000 square kilometers, and calculate the net magnetic flux through the rest of the Earth's surface (the entire surface excluding Arizona). Is the flux outward or inward?

23. The magnetic dipole moment of the Earth is 8.0×10^{22} J/T. (*a*) If the origin of this magnetism were a magnetized iron sphere at the center of the Earth, what would be its radius? (*b*) What fraction of the volume of the Earth would the sphere occupy? The density of the Earth's inner core is 14 g/cm^3. The magnetic dipole moment of an iron atom is 2.1×10^{-23} J/T.

24. Use the results displayed in Problem 9 to predict the value of the Earth's magnetic field (magnitude and inclination) at (*a*) the magnetic equator; (*b*) a point at magnetic latitude 60°; and (*c*) the north magnetic pole.

25. Find the altitude above the Earth's surface where the Earth's magnetic field has a magnitude one-half the surface value at the same magnetic latitude. (Use the dipole field approximation given in Problem 9.)

26. Using the dipole field approximation to the Earth's magnetic field (see Problem 9), calculate the maximum strength of the magnetic field at the core-mantle boundary, which is 2900 km below the Earth's surface.

27. Use the properties of the dipole field displayed in Problem 9 to calculate the magnitude and inclination angle of the Earth's magnetic field at the north geographic pole. (Hint: The angle between the magnetic axis and the rotational axis of the Earth is 11.5°.) Why do the calculated values probably not agree with the measured values?

35-7 Gauss' Law for Magnetism

28. A Gaussian surface in the shape of a right circular cylinder has a radius of 13 cm and a length of 80 cm. Through one end there is an inward magnetic flux of 25 μWb. At the other end there is a uniform magnetic field of 1.6 mT, normal to the surface and directed outward. Calculate the net magnetic flux through the curved surface.

29. The magnetic flux through each of five faces of one member of a pair of dice is given by $\Phi_B = \pm N$ Wb, where N $(=1$ to $5)$ is the number of spots on the face. The flux is positive (outward) for N even and negative (inward) for N odd. What is the flux through the sixth face?

30. Figure 35-27 shows four arrangements of pairs of small compass needles, set up in a space in which there is no external magnetic field. Identify the equilibrium in each case as stable or unstable. For each pair consider only the torque acting on one needle due to the magnetic field set up by the other. Explain your answers.

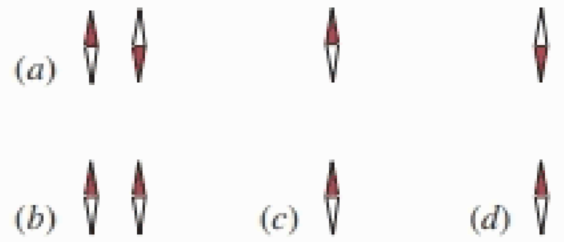


FIGURE 35-27. Exercise 30.

31. Two wires, parallel to the z axis and a distance $4r$ apart, carry equal currents i in opposite directions, as shown in Fig. 35-28. A circular cylinder of radius r and length L has its axis on the z axis, midway between the wires. Use Gauss' law for magnetism to calculate the net outward magnetic flux through the half of the cylindrical surface above the x axis. (Hint: Find the flux through that portion of the xz plane that is within the cylinder.)

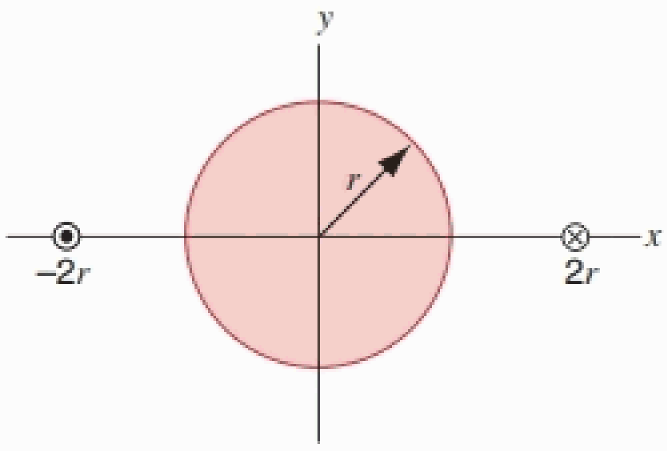


FIGURE 35-28. Exercise 31.

PROBLEMS

1. A thin, plastic disk of radius R has a charge q uniformly distributed over its surface. If the disk rotates at an angular frequency ω about its axis, show that magnetic dipole moment of the disk is

$$\mu = \frac{\omega q R^2}{4}.$$

(Hint: The rotating disk is equivalent to an array of current loops.)

2. (*a*) Calculate the magnetic moment of a uniformly charged, rotating sphere. (*b*) Show that the magnetic moment can be written as $\mu = qL/2m$ where L is the angular momentum of the sphere and m is the mass. (*c*) Show that this is not a good model for the structure of an electron. (Hint: The uniformly charged sphere must be divided into infinitesimal current loops and an expression for the magnetic moment found by integration.)

3. An electron with kinetic energy K_e travels in a circular path that is perpendicular to a uniform magnetic field, subject only to the force of the field. (*a*) Show that the magnetic dipole moment due to its orbital motion has magnitude $\mu = K_e/B$ and that it is in the direction opposite to that of $\vec{\mathbf{B}}$. (*b*) What is the magnitude and direction of the magnetic dipole moment of a positive ion with kinetic energy K_i under the same circumstances? (*c*) An ionized gas consists of 5.28×10^{21} electrons/m^3 and the same number of ions/m^3. Take the average electron kinetic energy to be 6.21×10^{-20} J and the average ion kinetic energy to be 7.58×10^{-21} J. Calculate the magnetization of the gas for a magnetic field of 1.18 T.

4. A paramagnetic substance is (weakly) attracted to a pole of a magnet. Figure 35-29 shows a model of this phenomenon. The "paramagnetic substance" is a current loop L, which is placed on the axis of a bar magnet nearer to its north pole than its south pole. Because of the torque $\vec{\tau} = \vec{\mu} \times \vec{\mathbf{B}}$ exerted on the loop by the $\vec{\mathbf{B}}$ field of the bar magnet, the magnetic dipole moment $\vec{\mu}$ of the loop will align itself to be parallel to $\vec{\mathbf{B}}$. (*a*) Make a sketch showing the $\vec{\mathbf{B}}$ field lines due to the bar magnet. (*b*) Show the direction of the current i in the loop. (*c*) Using $d\vec{\mathbf{F}}_B = i\, d\vec{\mathbf{s}} \times \vec{\mathbf{B}}$, show from (*a*) and (*b*) that the net force on L is toward the north pole of the bar magnet.

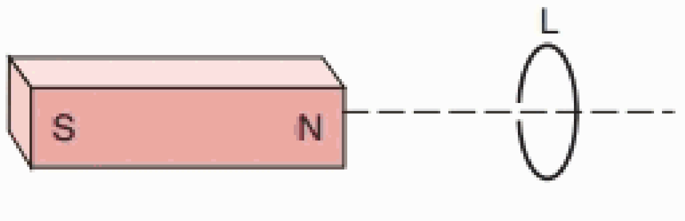


FIGURE 35-29. Problems 4 and 5.

5. A diamagnetic substance is (weakly) repelled by a pole of a magnet. Figure 35-29 shows a model of this phenomenon. The "diamagnetic substance" is a current loop L that is placed on the axis of a bar magnet nearer to its north pole than its south pole. Because the substance is diamagnetic, the magnetic moment $\vec{\mu}$ of the loop will align itself to be antiparallel to the $\vec{\mathbf{B}}$ field of the bar magnet. (*a*) Make a sketch showing the $\vec{\mathbf{B}}$ field lines due to the bar magnet. (*b*) Show the direction of the current i in the loop. (*c*) Using $d\vec{\mathbf{F}}_B = i\,d\vec{\mathbf{s}} \times \vec{\mathbf{B}}$, show from (*a*) and (*b*) that the net force on L is away from the north pole of the bar magnet.

6. Consider a solid containing N atoms per unit volume, each atom having a magnetic dipole moment $\vec{\mu}$. Suppose the direction of $\vec{\mu}$ can be only parallel or antiparallel to an externally applied magnetic field $\vec{\mathbf{B}}$ (this will be the case if $\vec{\mu}$ is due to the spin of a single electron). According to statistical mechanics, it can be shown that the probability of an atom being in a state with energy U is proportional to $e^{-U/kT}$, where T is the temperature and k is the Boltzmann constant. Thus, since $U = -\vec{\mu}\cdot\vec{\mathbf{B}}$, the fraction of atoms whose dipole moment is parallel to $\vec{\mathbf{B}}$ is proportional to $e^{\mu B/kT}$ and the fraction of atoms whose dipole moment is antiparallel to $\vec{\mathbf{B}}$ is proportional to $e^{-\mu B/kT}$. (*a*) Show that the magnetization of this solid is $M = N\mu \tanh(\mu B/kT)$. Here tanh is the hyperbolic tangent function: $\tanh x = (e^x - e^{-x})/(e^x + e^{-x})$. (*b*) Show that (*a*) reduces to $M = N\mu^2 B/kT$ for $\mu B \ll kT$. (*c*) Show that (*a*) reduces to $M = N\mu$ for $\mu B \gg kT$. (*d*) Show that (*b*) and (*c*) agree qualitatively with Fig. 35-10.

7. Consider an atom in which an electron moves in a circular orbit with radius r and angular frequency ω_0. A magnetic field is applied perpendicular to the plane of the orbit. As a result of the magnetic force, the electron circulates in an orbit with the same radius r but with a new angular frequency $\omega = \omega_0 + \Delta\omega$. (*a*) Show that, when the field is applied, the change in the centripetal acceleration of the electron is $2r\omega_0\,\Delta\omega$. (*b*) Assuming that the change in centripetal acceleration is entirely due to the magnetic force, derive Eq. 35-20.

8. The Earth has a magnetic dipole moment of 8.0×10^{22} J/T. (*a*) What current would have to be set up in a single turn of wire going around the Earth at its magnetic equator if we wished to set up such a dipole? (*b*) Could such an arrangement be used to cancel out the Earth's magnetism at points in space well above the Earth's surface? (*c*) On the Earth's surface?

9. The magnetic field of the Earth can be approximated as a dipole magnetic field, with horizontal and vertical components, at a point a distance r from the Earth's center, given by

$$B_\text{h} = \frac{\mu_0\mu}{4\pi r^3}\cos L_\text{m}, \qquad B_\text{v} = \frac{\mu_0\mu}{2\pi r^3}\sin L_\text{m},$$

where L_m is the magnetic latitude (latitude measured from the magnetic equator toward the north or south magnetic pole). The magnetic dipole moment μ is $8.0 \times 10^{22}\ \text{A}\cdot\text{m}^2$. (*a*) Show that the strength at latitude L_m is given by

$$B = \frac{\mu_0\mu}{4\pi r^3}\sqrt{1 + 3\sin^2 L_\text{m}}.$$

(*b*) Show that the inclination ϕ_i of the magnetic field is related to the magnetic latitude L_m by

$$\tan\phi_\text{i} = 2\tan L_\text{m}.$$

CHAPTER 36

INDUCTANCE

In Chapter 30 we studied the behavior of capacitors, which accumulate charge and thus set up an electric field in which energy is stored. In this chapter we study a device called an inductor, *in which energy is stored in the magnetic field surrounding its current-carrying wires.*

We determine the inductance of an inductor through the application of Faraday's law, in which a changing current produces an emf. The inductance measures the ability of the inductor to store energy in the magnetic field of a current. In a circuit, an inductor provides "inertia" that opposes a change in current. We will also analyze circuits containing only a capacitor and an inductor, in which the stored energy can oscillate back and forth between the two elements.

36-1 INDUCTANCE

An *inductor* is a circuit element that stores energy in the magnetic field surrounding its current-carrying wires, just as a capacitor stores energy in the electric field between its charged plates. An inductor is characterized by its *inductance,* which depends on its geometrical properties; in a similar way, in Chapter 30 we characterized a capacitor by its capacitance, which also depends on its geometrical properties.

Figure 36-1 shows an inductor, such as an ideal solenoid, carrying a current i that produces a magnetic field $\vec{\mathbf{B}}$ in its interior. If we change the current, thereby changing $\vec{\mathbf{B}}$ and the magnetic flux through the solenoid, Faraday's law shows that there is an emf generated in the inductor. The inductance L is defined* to be the proportionality constant that relates the rate of change of current to the induced emf:

$$\mathcal{E}_L = L\frac{di}{dt}. \tag{36-1}$$

This equation is similar to the defining equation for capacitance ($\Delta V_C = q/C$). Like the capacitance, the inductance is always taken to be a positive quantity.

Equation 36-1 shows that the SI unit of inductance is the volt · second/ampere. This combination of units has been given the special name of the *henry* (abbreviation H), so that

$$1 \text{ henry} = 1 \text{ volt} \cdot \text{second/ampere}.$$

This unit is named after Joseph Henry (1797–1878), an American physicist and a contemporary of Faraday. In an electrical circuit diagram, an inductor is represented by the symbol , which resembles the shape of a solenoid.

To find the relationship between the sign of $\mathcal{E}_L$ and the sign of di/dt, we use Lenz' law. Let us decrease the current i in the solenoid of Fig. 36-1. This decrease is the change that, according to Lenz' law, the inductance must oppose.

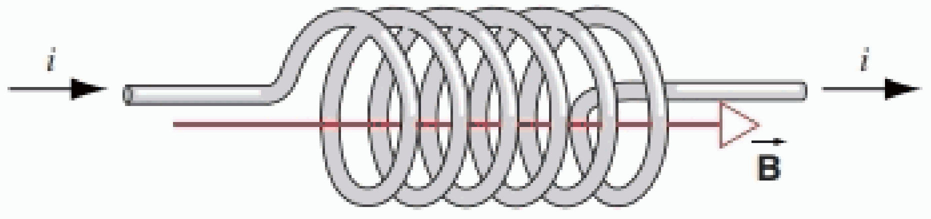


FIGURE 36-1. An arbitrary inductor, represented as a solenoid. The current i establishes a magnetic field $\vec{\mathbf{B}}$.

---

*Strictly speaking, L is called the "self inductance," which means that the change in current in the device itself causes the emf across the device. A similar quantity associated with two nearby current-carrying elements is the "mutual inductance," in which a change in current in one element causes an emf in the other.

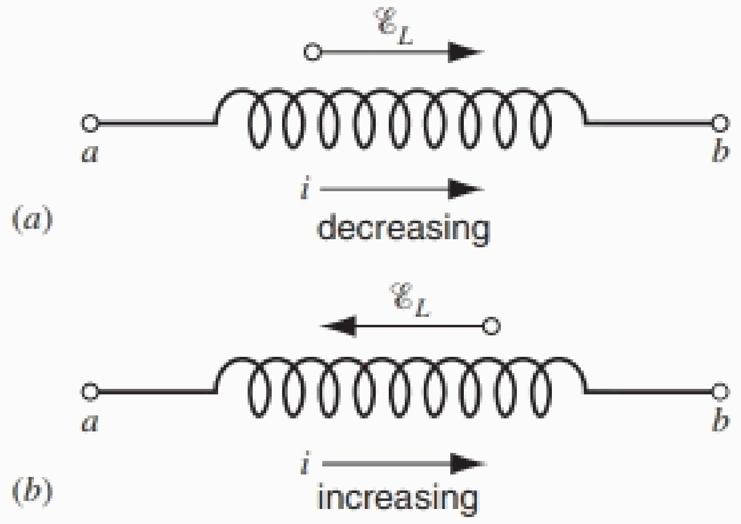


FIGURE 36-2. (*a*) A *decreasing* current induces in the inductor an emf that opposes the decrease in current. (*b*) An *increasing* current induces in the inductor an emf that opposes the increase.

To oppose the falling current, the induced emf must provide an additional current in the *same* direction as i.

If instead we increase the current, Lenz' law shows that the increase is opposed by the inductance through an additional current in a direction *opposite* to i.

In each case, the induced emf acts to oppose the *change* in the current. Figure 36-2 summarizes the relationship between the sign of di/dt and the sign of $\mathcal{E}_L$. In Fig. 36-2*a*, the induced potential differences are such that V_b is greater than V_a, so that $V_b - V_a = |L\, di/dt|$. Since i is decreasing, di/dt is negative, so we can write this as

$$V_b - V_a = -L\, di/dt. \qquad (36\text{-}2)$$

In Fig. 36-2*b*, di/dt is positive and V_a is greater than V_b, so Eq. 36-2 applies in this case as well. Equation 36-2 is particularly useful when we use the loop theorem to analyze circuits containing inductors.

36-2 CALCULATING THE INDUCTANCE

We can use Eq. 36-1 to find the inductance of an inductor of any particular size or shape. Our method will be to use the magnetic field of the current-carrying inductor to find the flux through the device, and then to use Faraday's law to find the emf corresponding to a changing current. Equation 36-1 then gives the inductance. We used a similar method to find the capacitance in Section 30-3 by calculating the potential difference between the two charged plates and then using the proportionality between ΔV_C and q to find C.

Suppose the current in the inductor sets up a magnetic field $\vec{\mathbf{B}}$, which we can calculate from the size and shape of the inductor and from the distribution of current. This enables the magnetic flux Φ_B through each turn of the coil to be obtained. For the cases we will consider, the flux has the same value for each of the N turns of the coil, so that the total flux is $N\Phi_B$, This quantity is known as the number of *flux linkages* of the inductor.

The emf can be found from Faraday's law,

$$\mathcal{E}_L = -\frac{d(N\Phi_B)}{dt}. \qquad (36\text{-}3)$$

Equations 36-1 and 36-3 relate the emf in an inductor to the derivative of the current (Eq. 36-1) or to the derivative of a property that is proportional to the current (Φ_B in Eq. 36-3). Comparing the two equations (and taking the magnitude of all quantities), we find

$$L\frac{di}{dt} = \frac{d(N\Phi_B)}{dt}.$$

Integrating with respect to the time (and assuming that $\Phi_B = 0$ when $i = 0$), we find

$$Li = N\Phi_B,$$

or

$$L = \frac{N\Phi_B}{i}. \qquad (36\text{-}4)$$

Equation 36-4, which is based on Faraday's law, permits the inductance to be found directly from the number of flux linkages. Note that, since Φ_B is proportional to the current i, the ratio in Eq. 36-4 is *independent* of i and thus the inductance (like the capacitance) depends only on the geometry of the device.

The Inductance of a Solenoid

Let us apply Eq. 36-4 to calculate L for a section of length l of a long solenoid of cross-sectional area A; we assume the section is near the center of the solenoid so that edge effects need not be considered. In Section 33-5, the magnetic field B inside a solenoid carrying a current i was shown to be

$$B = \mu_0 ni, \qquad (36\text{-}5)$$

where n is the number of turns per unit length. The number of flux linkages in the length l is

$$N\Phi_B = (nl)(BA),$$

which becomes, after substituting for B,

$$N\Phi_B = \mu_0 n^2 liA. \qquad (36\text{-}6)$$

Equation 36-4 then gives the inductance directly:

$$L = \frac{N\Phi_B}{i} = \frac{\mu_0 n^2 liA}{i} = \mu_0 n^2 lA. \qquad (36\text{-}7)$$

The inductance per unit length of the solenoid can be written

$$\frac{L}{l} = \mu_0 n^2 A. \qquad (36\text{-}8)$$

This expression involves only geometrical factors—the cross-sectional area and the number of turns per unit length. The inductance does *not* depend on the current or the magnetic field. The proportionality to n^2 is expected; if we double the number of turns per unit length, not only is the number N of turns doubled, but the flux Φ_B through *each turn* is doubled, and the number of flux linkages increases by a factor of 4, as does the inductance.

Equations 36-7 and 36-8 are valid for a solenoid of length very much greater than its radius. We have neglected the spreading of the magnetic field lines near the end of a solenoid, just as we neglected the fringing of the electric field near the edges of the plates of a capacitor.

The Inductance of a Toroid

We now calculate the inductance of a toroid of rectangular cross section, as shown in Fig. 36-3. The magnetic field B in a toroid was given by Eq. 33-36:

$$B = \frac{\mu_0 i N}{2\pi r}, \tag{36-9}$$

where N is the total number of turns of the toroid. Note that the magnetic field is not constant inside the toroid but varies with the radius r.

The flux Φ_B through the cross section of the toroid is

$$\Phi_B = \int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = \int_a^b B(h\,dr) = \int_a^b \frac{\mu_0 i N}{2\pi r} h\,dr$$
$$= \frac{\mu_0 i N h}{2\pi} \int_a^b \frac{dr}{r} = \frac{\mu_0 i N h}{2\pi} \ln \frac{b}{a},$$

where $h\,dr$ is the area of the elementary strip of width dr shown in Fig. 36-3. The inductance can then be found directly from Eq. 36-4:

$$L = \frac{N\Phi_B}{i} = \frac{\mu_0 N^2 h}{2\pi} \ln \frac{b}{a}. \tag{36-10}$$

Once again, L depends only on geometrical factors.

Inductors with Magnetic Materials

In Section 30-6 we showed that filling a capacitor with a dielectric substance increases the capacitance. This enables the capacitor to store more charge on its plates or more energy in its electric field. In a similar way, filling an inductor with a magnetic material can increase the inductance.

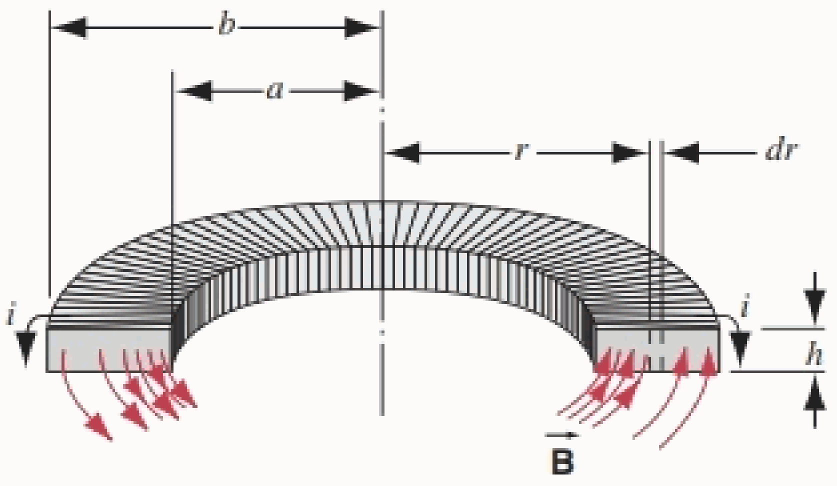


FIGURE 36-3. A cross section of a toroid, showing the current in the windings and the magnetic field in the interior.

Equation 35-16 shows that the presence of a magnetic material changes the magnetic field in the material from the applied field $\vec{\mathbf{B}}_0$ to $\vec{\mathbf{B}} = \kappa_m \vec{\mathbf{B}}_0$, where κ_m is the permeability of the material. The magnetic field in the interior of the inductor is contained in the factor Φ_B in Eq. 36-4. If the magnetic field in the solenoid is everywhere multiplied by the factor κ_m, then Φ_B will also be multiplied by that factor, and we conclude that

$$L = \kappa_m L_0, \tag{36-11}$$

where L is the inductance of the inductor with the magnetic material present and L_0 is the inductance of the empty inductor.

Because the permeabilities of paramagnetic or diamagnetic substances do not differ substantially from 1, the inductances of inductors filled with such substances are nearly equal to their values when empty, and no major change in the properties of the inductor is obtained by filling the inductor with a paramagnetic or a diamagnetic material. In the case of a ferromagnetic material, however, substantial changes can occur. Although the permeability is not defined in general for ferromagnetic materials (because the total field does not increase in linear proportion to the applied field), under particular circumstances B can be several thousand times B_0. Thus the "effective" permeability for a ferromagnet can have values in the range of 10^3 to 10^4, and the inductance of an inductor filled with ferromagnetic material (that is, one in which the windings are made on a core of a material such as iron) can be greater than the inductance of a similar set of windings on an empty core by a factor of 10^3 to 10^4. Ferromagnetic cores provide the means to obtain large inductances, just as dielectric materials in capacitors permit large capacitances to be obtained.

SAMPLE PROBLEM 36-1. A section of a solenoid of length $l = 12$ cm and having a circular cross section of diameter $d = 1.6$ cm carries a steady current of $i = 3.80$ A. The section contains 75 turns along its length. (*a*) What is the inductance of the solenoid when the core is empty? (*b*) The current is reduced at a constant rate to 3.20 A in a time of 15 s. What is the resulting emf developed by the solenoid, and in what direction does it act?

Solution (*a*) The inductance of the solenoid is found from Eq. 36-7:

$$\begin{aligned} L &= \mu_0 n^2 l A \\ &= (4\pi \times 10^{-7}\ \mathrm{H/m})(75\ \mathrm{turns}/0.12\ \mathrm{m})^2(0.12\ \mathrm{m})(\pi)(0.008\ \mathrm{m})^2 \\ &= 1.2 \times 10^{-5}\ \mathrm{H} = 12\ \mu\mathrm{H}. \end{aligned}$$

Note that we have expressed μ_0 in units of H/m. An inductance can always be expressed as μ_0 times a quantity with the dimension of length. A similar situation holds for capacitance; see Section 30-3.

(*b*) The rate at which the current changes is

$$\frac{di}{dt} = \frac{3.20\ \mathrm{A} - 3.80\ \mathrm{A}}{15\ \mathrm{s}} = -0.040\ \mathrm{A/s},$$

and the corresponding emf has magnitude given by Eq. 36-1:

$$\mathcal{E}_L = |L\,di/dt| = (12\ \mu\mathrm{H})(0.040\ \mathrm{A/s}) = 0.48\ \mu\mathrm{V}.$$

Because the current is decreasing, the induced emf must act in the same direction as the current, so that the induced emf opposes the decreases in the current.

SAMPLE PROBLEM 36-2. The core of the solenoid of Sample Problem 36-1 is filled with iron while the current is held constant at 3.20 A. The magnetization of the iron is saturated such that $B = 1.4$ T. What is the resulting inductance?

Solution The "effective" permeability of the core subject to this particular applied field is determined from

$$\kappa_m = \frac{B}{B_0} = \frac{B}{\mu_0 n i}$$

$$= \frac{1.4\ \mathrm{T}}{(4\pi \times 10^{-7}\ \mathrm{T \cdot m/A})(75\ \mathrm{turns}/0.12\ \mathrm{m})(3.20\ \mathrm{A})} = 557.$$

The inductance is given by Eq. 36-11 as

$$L = \kappa_m L_0 = (557)(12\ \mu\mathrm{H}) = 6.7\ \mathrm{mH}.$$

36-3 *LR* CIRCUITS

In this section we consider the behavior of circuits containing a resistor and an inductor in series. In several ways, this topic is similar to the analysis of *RC* circuits, which we discussed in Section 31-7. There we found that when a battery is connected to a series combination of a resistor and a capacitor, the charge on the capacitor approaches its maximum exponentially with a time constant $\tau_C = RC$. Similarly, the discharge of a capacitor through a resistor is also exponential with the same time constant.

Figure 36-4 shows a circuit in which a resistor R and an inductor L are connected in series. Suitable devices are available for measuring voltage differences across the resistor (ΔV_R) and inductor (ΔV_L). A switch S can connect a battery of emf $\mathcal{E}$ into the circuit. Initially, no current flows in the circuit. When the switch is moved to a, the current in the resistor starts to rise. If the inductor were not present, the current would quickly rise to a steady value $\mathcal{E}/R$. The inductor, however, gives an induced emf $\mathcal{E}_L$, which, according to Lenz' law, opposes the rise in current. That is, it opposes the polarity of the battery emf. The current in the circuit depends on two emfs: a constant emf $\mathcal{E}$ due to the battery and a variable emf $\mathcal{E}_L$ of the opposite sign due to the inductance. As long as this second emf is present, the current in the circuit is smaller than $\mathcal{E}/R$.

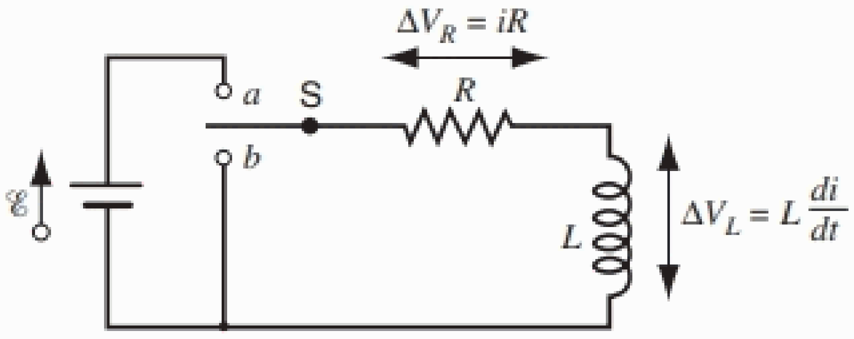


FIGURE 36-4. An *LR* circuit.

As time goes on, the current increases less rapidly, and the induced emf, which is proportional to di/dt, becomes smaller. The more the current increases, the slower its increase becomes, and the induced emf becomes correspondingly smaller. As the induced emf becomes negligibly small, the current approaches its maximum value $\mathcal{E}/R$.

In a circuit such as this, we can consider the inductor to behave like an infinite resistance just after the battery is switched into the circuit. Much later the inductor behaves like a resistance of zero as the current approaches its steady value.

Now let us analyze this circuit quantitatively. When the switch is at a, a clockwise current is set up in the circuit and the loop rule (Section 31-3) gives

$$\mathcal{E} - iR - L\frac{di}{dt} = 0$$

or

$$\mathcal{E} = iR + L\frac{di}{dt}, \tag{36-12}$$

where we have used Eq. 36-2 for the potential difference across the inductor.

The solution to Eq. 36-12 is a function $i(t)$ chosen so that when it and its first derivative are substituted into Eq. 36-12 the equation is satisfied. Equation 36-12 has exactly the same form as Eq. 31-25 for the *RC* circuit, and it should not be surprising that its solution has the same form as well (see Eq. 31-27):

$$i(t) = \frac{\mathcal{E}}{R}(1 - e^{-t/\tau_L}), \tag{36-13}$$

where

$$\tau_L = \frac{L}{R}. \tag{36-14}$$

The *inductive time constant* τ_L indicates how rapidly the current approaches its steady value, in analogy with the capacitative time constant τ_C. Note that Eq. 36-13 gives $i = 0$ at $t = 0$ and $i \rightarrow \mathcal{E}/R$ as $t \rightarrow \infty$, as we expect.

By taking the derivative of Eq. 36-13 and substituting i and di/dt into Eq. 36-12, you should verify that Eq. 36-13 is indeed the solution of Eq. 36-12.

To show that the quantity $\tau_L = L/R$ has the dimension of time, we have

$$[\tau_L] = \frac{[L]}{[R]} = \frac{\text{henry}}{\text{ohm}} = \frac{\text{volt} \cdot \text{second/ampere}}{\text{ohm}}$$

$$= \left(\frac{\text{volt}}{\text{ampere} \cdot \text{ohm}}\right) \text{second} = \text{second},$$

where the quantity in parentheses equals 1 because 1 ohm = 1 volt/ampere (as in $R = V/i$).

The physical significance of τ_L follows from Eq. 36-13. If we put $t = \tau_L$ into this equation, it reduces to

$$i = \frac{\mathscr{E}}{R}(1 - e^{-1}) = (1 - 0.37)\frac{\mathscr{E}}{R} = 0.63\frac{\mathscr{E}}{R}.$$

The time constant τ_L is that time at which the current in the circuit is less than its final steady value $\mathscr{E}/R$ by a factor of $1/e$ (about 37%).

Figure 36-5 shows the potential difference ΔV_R $[= i(t)R]$ across the resistor R and the potential difference ΔV_L $[= L(di/dt)]$ across the ideal inductor. From Eq. 36-13 we obtain

$$\Delta V_R = iR = \mathscr{E}(1 - e^{-t/\tau_L}) \quad \text{and} \quad \Delta V_L = L\frac{di}{dt} = \mathscr{E}e^{-t/\tau_L}, \tag{36-15}$$

which are the quantities plotted in Fig. 36-5. From Eqs. 36-15 we obtain $\Delta V_R + \Delta V_L = \mathscr{E}$, which we could find by applying the loop rule to the circuit. This result also follows from adding the graphs of Figs. 36-5*a* and 36-5*b*.

If we throw the switch in Fig. 36-4 from *a* to *b* when the current has some value i_0, the effect is to remove the battery from the circuit. The loop theorem now gives

$$L\frac{di}{dt} + iR = 0. \tag{36-16}$$

By direct substitution or integration, we can find the solution of this equation to be

$$i(t) = i_0 e^{-t/\tau_L}, \tag{36-17}$$

where i_0 is the current at $t = 0$ (which now means the time that the switch is thrown from *a* to *b*). The decay of the current to zero occurs with the same exponential time constant $\tau_L = L/R$ as the rise in current.

Figure 36-6 shows the potential differences ΔV_R and ΔV_L across the resistor and inductor for the case in which the switch is left at *a* for a time that is very long compared with τ_L, so that we can regard the current as having reached its maximum value $\mathscr{E}/R$ when we throw the switch to *b*.

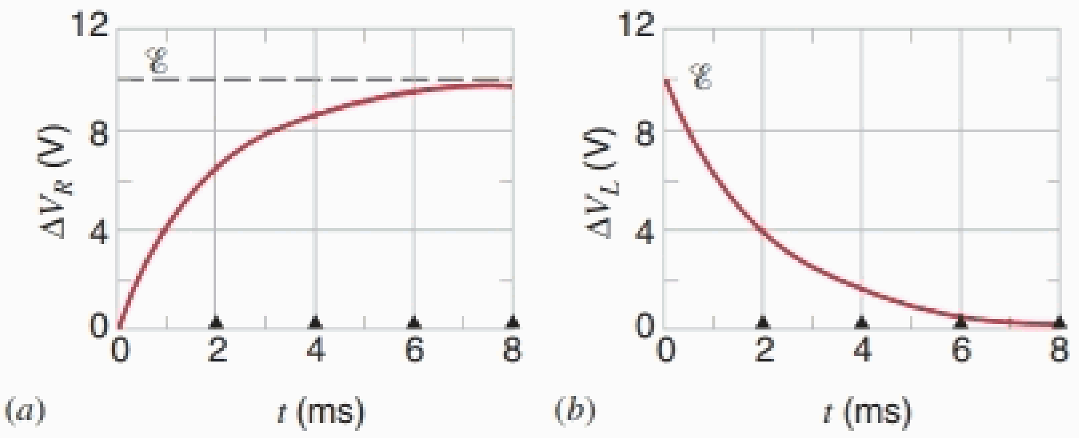


FIGURE 36-5. The variation with time of (*a*) ΔV_R, the potential difference across the resistor in the circuit of Fig. 36-4, and (*b*) ΔV_L, the potential difference across the inductor in that circuit. The curves are drawn for $R = 2000\ \Omega$, $L = 4.0$ H, and $\mathscr{E} = 10$ V. The inductive time constant τ_L is 2 ms; successive intervals equal to τ_L are marked by the triangles along the horizontal axis.

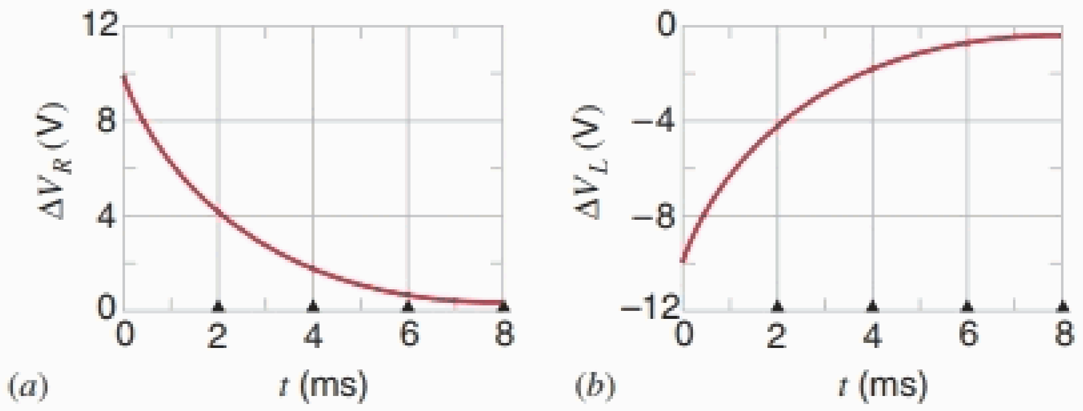


FIGURE 36-6. Same as Fig. 36-5 for the circuit of Fig. 36-4, with the switch thrown to *b* after it has been at *a* for a long time.

Using Eq. 36-17 to obtain ΔV_R and ΔV_L in a manner similar to Eq. 36-15, you should be able to show that $\Delta V_R + \Delta V_L = 0$, which you can also see directly by adding the graphs in Figs. 36-6*a* and 36-6*b*.

In constructing the circuit of Fig. 36-4, we must use a special kind of switch called a "make before break" switch. This switch completes the connection at *b* before it breaks the connection at *a*. If we did not use this type of switch, the current would try to continue flowing with the switch terminal between *a* and *b* by jumping the gap from *a* to the switch terminal with a spark that could contain all the energy stored in the inductor.

SAMPLE PROBLEM 36-3. A solenoid has an inductance of 53 mH and a resistance of 0.37 Ω. If it is connected to a battery, how long will it take for the current to reach one-half its final steady-state value?

Solution The steady-state value of the current, which is reached at $t \to \infty$, is $\mathscr{E}/R$ from Eq. 36-13. If the current has half this value at a particular time t, this equation becomes

$$\frac{1}{2}\frac{\mathscr{E}}{R} = \frac{\mathscr{E}}{R}(1 - e^{-t/\tau_L}),$$

or

$$e^{-t/\tau_L} = \tfrac{1}{2}.$$

Solving for t by rearranging and taking the (natural) logarithm of each side, we find

$$t = \tau_L \ln 2 = \frac{L}{R}\ln 2 = \frac{53 \times 10^{-3}\ \text{H}}{0.37\ \Omega}\ln 2 = 0.10\ \text{s}.$$

36-4 ENERGY STORAGE IN A MAGNETIC FIELD

In Chapter 28 we learned that a collection of electric charges can be described by its potential energy. Equivalently, we can say that the energy is stored in the electric field due to the charges. For example, the work done in separating two charges of opposite sign is stored in the energy of the electric field of the charges; that energy can be recovered by allowing the charges to move together again.

In a similar manner, there is energy stored in the region around a current-carrying wire where a magnetic field exists. For example, two wires carrying parallel currents attract each other; the work done in separating the wires is stored in the energy of the magnetic field of the wires, and we could recover that energy by allowing the wires to move back together.

In Chapter 30 we discussed the energy stored in a capacitor. We generalized that discussion by regarding the energy as stored in the electric field of the capacitor, and by inference we obtained the energy stored by *any* electric field. Here we consider the energy stored in an inductor, which will lead us to a calculation of the energy stored in *any* magnetic field.

Let us return to the circuit of Fig. 36-4, with the switch at a. Using the loop rule (Section 31-3), we obtain Eq. 36-12: $\mathscr{E} = iR + L\,di/dt$. We recall from Chapter 31 that the loop rule is essentially a statement of conservation of energy in the circuit. We can see this more clearly by multiplying both sides of Eq. 36-12 by the current i:

$$\mathscr{E}i = i^2R + Li\frac{di}{dt}, \tag{36-18}$$

which has the following physical interpretation in terms of work and energy:

1. If a charge dq passes through the battery in Fig. 36-4 in a time dt, the battery does work on it in the amount $\mathscr{E}\,dq$. The *rate* of doing work is $(\mathscr{E}\,dq)/dt$ or $\mathscr{E}i$. Thus the left side of Eq. 36-18 is the *rate at which the emf device delivers energy to the circuit.*

2. The second term in Eq. 36-18, i^2R, is the *rate at which energy is dissipated in the resistor.* This energy appears as the internal energy associated with atomic motions in the resistor.

3. Energy delivered to the circuit but not dissipated in the resistor must, by our hypothesis, be stored in the magnetic field. Since Eq. 36-18 represents a statement of the conservation of energy for LR circuits, the last term must represent the *rate at which energy is stored in the magnetic field.*

Let U_B represent the energy stored in the magnetic field; then the rate at which energy is stored is dU_B/dt. Equating the rate of energy storage to the last term of Eq. 36-18, we obtain

$$\frac{dU_B}{dt} = Li\frac{di}{dt} \tag{36-19}$$

or

$$dU_B = Li\,di. \tag{36-20}$$

Suppose we start with no current in the inductor ($i = 0$) and no stored energy in its magnetic field. We gradually increase the current to the final value i. The energy U_B stored in the magnetic field can be found by integrating Eq. 36-20, which gives

$$\int_0^{U_B} dU_B = \int_0^i Li\,di$$

or

$$U_B = \tfrac{1}{2}Li^2, \tag{36-21}$$

which represents the total stored magnetic energy in an inductance L carrying a current i.

If the switch in Fig. 36-4 is thrown from a to b after a current i_0 is established, the stored energy in the inductor dissipates through Joule heating in the resistor.

An analogous situation holds in charging and discharging a capacitor. When the capacitor has accumulated a charge q, the energy stored in the electric field is

$$U_E = \frac{1}{2}\frac{q^2}{C}.$$

We derived this expression in Section 30-5 by setting the stored energy equal to the work that must be done in setting up the field. The capacitor can discharge through a resistor, in which case the stored energy is again dissipated through Joule heating.

SAMPLE PROBLEM 36-4. A coil has an inductance of 53 mH and resistance of 0.35 Ω. (*a*) If a 12-V emf is applied, how much energy is stored in the magnetic field after the current has built up to its maximum value? (*b*) In terms of τ_L, how long does it take for the stored energy to reach half of its maximum value?

Solution (*a*) From Eq. 36-13 the maximum current is

$$i_m = \frac{\mathscr{E}}{R} = \frac{12\text{ V}}{0.35\ \Omega} = 34.3\text{ A}.$$

Substituting this current into Eq. 36-21, we find the stored energy:

$$U_B = \tfrac{1}{2}Li_m^2 = \tfrac{1}{2}(53\times10^{-3}\text{ H})(34.3\text{ A})^2$$
$$= 31\text{ J}.$$

(*b*) Let i be the current at the instant the stored energy has half its maximum value. Then

$$\tfrac{1}{2}Li^2 = (\tfrac{1}{2})\tfrac{1}{2}Li_m^2$$

or

$$i = i_m/\sqrt{2}.$$

However, i is given by Eq. 36-13 and i_m (see above) is $\mathscr{E}/R$, so that

$$\frac{\mathscr{E}}{R}(1 - e^{-t/\tau_L}) = \frac{\mathscr{E}}{\sqrt{2}R}.$$

This can be written as

$$e^{-t/\tau_L} = 1 - 1/\sqrt{2} = 0.293,$$

which yields

$$-\frac{t}{\tau_L} = \ln 0.293 = -1.23$$

or

$$t = 1.23\tau_L.$$

The stored energy reaches half its maximum value after 1.23 time constants.

SAMPLE PROBLEM 36-5. A 3.56-H inductor is placed in series with a 12.8-Ω resistor, and an emf of 3.25 V is suddenly applied to the combination. At 0.278 s (which is one inductive time constant) after the contact is made, find (*a*) the rate P at which energy is being delivered by the battery, (*b*) the rate P_R at which internal energy appears in the resistor, and (*c*) the rate P_B at which energy is stored in the magnetic field.

Solution (*a*) The current is given by Eq. 36-13. At $t = \tau_L$, we obtain

$$i = \frac{\mathcal{E}}{R}(1 - e^{-t/\tau_L}) = \frac{3.25\ \text{V}}{12.8\ \Omega}(1 - e^{-1}) = 0.1605\ \text{A}.$$

The rate P at which the battery delivers energy is then

$$P = \mathcal{E}i = (3.25\ \text{V})(0.1605\ \text{A}) = 0.522\ \text{W}.$$

(*b*) The rate P_R at which energy is dissipated in the resistor is given by

$$P_R = i^2R = (0.1605\ \text{A})^2(12.8\ \Omega) = 0.330\ \text{W}.$$

(*c*) The rate P_B $(= dU_B/dt)$ at which energy is being stored in the magnetic field in given by Eq. 36-19. Differentiating Eq. 36-13 and using $t = \tau_L = L/R$, we obtain

$$\frac{di}{dt} = \frac{\mathcal{E}}{L}e^{-t/\tau_L} = \frac{3.25\ \text{V}}{3.56\ \text{H}}e^{-1} = 0.3358\ \text{A/s}.$$

From Eq. 36-19 the desired rate is then

$$P_B = \frac{dU_B}{dt} = Li\frac{di}{dt}$$

$$= (3.56\ \text{H})(0.1605\ \text{A})(0.3358\ \text{A/s}) = 0.192\ \text{W}.$$

Note that, as required by energy conservation,

$$P = P_R + P_B,$$

or

$$P = 0.330\ \text{W} + 0.192\ \text{W} = 0.522\ \text{W}.$$

Energy Density and the Magnetic Field

We now derive an expression for the *energy density* (energy per unit volume) u_B in a magnetic field. Consider a very long solenoid of cross-sectional area A whose interior contains no material. A portion of length l far from either end encloses a volume Al. The magnetic energy stored in this portion of the solenoid must lie entirely within this volume because the magnetic field outside the solenoid is essentially zero. Moreover, the stored energy must be uniformly distributed throughout the volume of the solenoid because the magnetic field is uniform everywhere inside. Thus we can write the energy density as

$$u_B = \frac{U_B}{Al}$$

or, since

$$U_B = \tfrac{1}{2}Li^2,$$

we have

$$u_B = \frac{\frac{1}{2}Li^2}{Al}.$$

in which L is the inductance of the solenoid of length l. To express this in terms of the magnetic field, we can solve Eq. 36-5 ($B = \mu_0 in$) for i and substitute in this equation. We can also substitute for L using the relation $L = \mu_0 n^2 lA$ (Eq. 36-7). Doing so yields finally

$$u_B = \frac{1}{2\mu_0}B^2. \tag{36-22}$$

This equation gives the energy density stored at any point (in a vacuum or in a nonmagnetic substance) where the magnetic field is $\vec{\mathbf{B}}$. The equation is true for all magnetic field configurations, even though we derived it by considering a special case, the solenoid. Equation 36-22 is to be compared with Eq. 30-28,

$$u_E = \tfrac{1}{2}\epsilon_0 E^2, \tag{36-23}$$

which gives the energy density (in a vacuum) at any point in an electric field. Note that both u_B and u_E are proportional to the square of the appropriate field quantity, B or E.

The solenoid plays a role for magnetic fields similar to that of the parallel-plate capacitor for electric fields. In each case we have a simple device that can be used for setting up a uniform field throughout a well-defined region of space and for deducing, in a simple way, properties of these fields.

SAMPLE PROBLEM 36-6. A long coaxial cable (Fig. 36-7) consists of two concentric cylindrical conductors with radii a and b, where $b \gg a$. Its central conductor carries a steady current i, and the outer conductor provides the return path. (*a*) Calculate the energy stored in the magnetic field for a length l of such a cable. (*b*) What is the inductance of a length l of the cable?

Solution (*a*) We assume that the inner conductor is so thin that we can neglect any magnetic energy stored in its interior. We

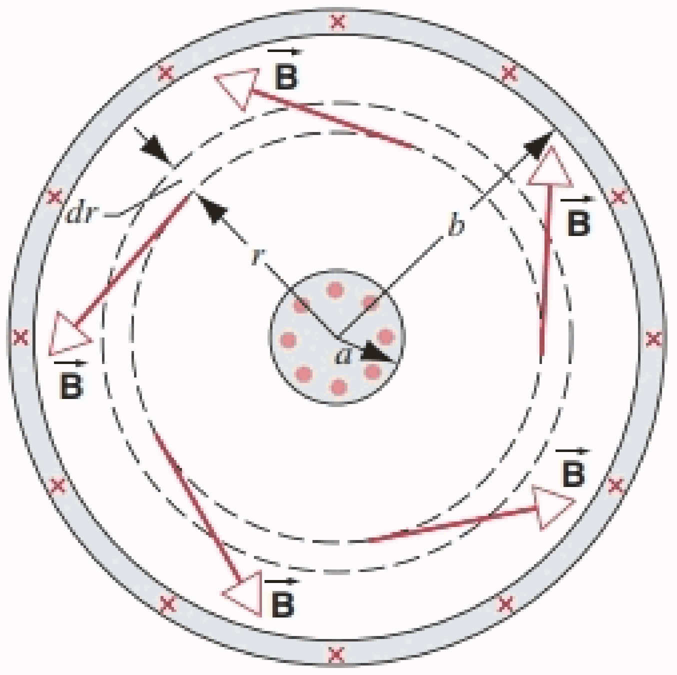


FIGURE 36-7. Sample Problem 36-6. Cross section of a coaxial cable, which carries steady equal but opposite currents in its inner and outer conductors. In the region between the conductors the lines of $\vec{\mathbf{B}}$ form circles.

make the same assumption for the outer conductor. If we construct an Ampèrian loop in the form of a circle of radius larger than the outer radius of the outer conductor, the net current passing through the surface bounded by the loop is zero (because the inner and outer conductors carry equal currents in opposite directions); we therefore conclude that $B = 0$ at all points beyond the outer conductor.

The magnetic energy is therefore confined to the region between the two conductors. The magnetic field in that region is identical to that of a straight wire, which we found in Sections 33-2 and 33-5. We can show this by constructing the circular Ampèrian loop shown in Fig. 36-7 and noting that the current in the outer conductor does not pass through the surface bounded by the loop and so does not contribute to Ampère's law. We therefore take $B = \mu_0 i/2\pi r$ (Eq. 33-13) in the region between the conductors.

The energy density for points between the conductors, from Eq. 36-22, is

$$u_B = \frac{1}{2\mu_0}B^2 = \frac{1}{2\mu_0}\left(\frac{\mu_0 i}{2\pi r}\right)^2 = \frac{\mu_0 i^2}{8\pi^2 r^2}.$$

Consider a volume element dV consisting of a cylindrical shell whose radii are r and $r + dr$ and whose length (perpendicular to the plane of Fig. 36-7) is l. The energy dU_B contained in it is

$$dU_B = u_B\,dV = \frac{\mu_0 i^2}{8\pi^2 r^2}(2\pi r l)(dr) = \frac{\mu_0 i^2 l}{4\pi}\frac{dr}{r}.$$

The total stored magnetic energy is found by integration:

$$U_B = \int dU_B = \frac{\mu_0 i^2 l}{4\pi}\int_a^b \frac{dr}{r} = \frac{\mu_0 i^2 l}{4\pi}\ln\frac{b}{a}.$$

(*b*) We can find the inductance L from Eq. 36-21 ($U_B = \frac{1}{2}Li^2$), which leads to

$$L = \frac{2U_B}{i^2} = \frac{\mu_0 l}{2\pi}\ln\frac{b}{a}.$$

You should also derive this expression directly from the definition of inductance, using the procedures of Section 36-2 (see Problem 5).

SAMPLE PROBLEM 36-7. Compare the energy required to set up, in a cube $l = 10$ cm on edge, (*a*) a uniform electric field of 1.0×10^5 V/m and (*b*) a uniform magnetic field of 1.0 T. Both these fields would be judged reasonably large but they are readily available in the laboratory.

Solution (*a*) In the electric case we have, where l^3 is the volume of the cube,

$$\begin{aligned} U_E &= u_E l^3 = \tfrac{1}{2}\epsilon_0 E^2 l^3 \\ &= (0.5)(8.9 \times 10^{-12}\ \mathrm{C^2/N\cdot m^2})(10^5\ \mathrm{V/m})^2(0.1\ \mathrm{m})^3 \\ &= 4.5 \times 10^{-5}\ \mathrm{J}. \end{aligned}$$

(*b*) In the magnetic case, from Eq. 36-22 we have

$$\begin{aligned} U_B = u_B l^3 = \frac{B^2}{2\mu_0}l^3 &= \frac{(1.0\ \mathrm{T})^2(0.1\ \mathrm{m})^3}{(2)(4\pi \times 10^{-7}\ \mathrm{T\cdot m/A})} \\ &= 400\ \mathrm{J}. \end{aligned}$$

In terms of fields normally available in the laboratory, much larger amounts of energy can be stored in a magnetic field than in an electric one, the ratio being about 10^7 in this example. Conversely, much more energy is required to set up a magnet field of reasonable laboratory magnitude than is required to set up an electric field of similarly reasonable magnitude.

36-5 ELECTROMAGNETIC OSCILLATIONS: QUALITATIVE

We now turn to a study of the properties of circuits that contain both a capacitor C and an inductor L. Such a circuit forms an *electromagnetic oscillator,* in which the current varies sinusoidally with time, much as the displacement of a mechanical oscillator varies with time. In fact, as we shall see, there are several analogies between electromagnetic and mechanical oscillators, which simplify our discussion of electromagnetic oscillations by allowing us to refer to our previous analysis of mechanical oscillations (Chapter 17).

For the time being, we assume the circuit to include no resistance. The circuit *with* resistance, which we consider in Section 36-7, is analogous to the damped oscillator we discussed in Section 17-7. We also assume that no source of emf is present in the circuit; oscillating circuits with emf present, which we also consider in Section 36-7, are analogous to forced mechanical oscillators such as those we discussed in Section 17-8.

With no source of emf present, the energy in the circuit comes from the energy initially stored in one or both of the components. Let us assume the capacitor C is charged (from some external source that does not concern us) so that it contains a charge q_m, at which time it is removed from the external source and connected to the inductor L. The LC circuit is shown in Fig. 36-8*a*. At first, the energy U_E stored in the capacitor is

$$U_E = \frac{1}{2}\frac{q_\mathrm{m}^2}{C}, \tag{36-24}$$

while the energy $U_B = \frac{1}{2}Li^2$ (Eq. 36-21) stored in the inductor is initially zero, because the current is zero.

The capacitor now starts to discharge through the inductor, positive charge carriers moving counterclockwise, as shown in Fig. 36-8*b*. A current $i = dq/dt$ now flows through the inductor, increasing its stored energy from zero. At the same time, the discharging of the capacitor reduces its stored energy. If the circuit is free of resistance, no energy is dissipated, and the decrease in the energy stored in the capacitor is exactly compensated by an increase in the energy stored in the inductor, such that the total energy remains constant. In effect, the electric field decreases and the magnetic field increases, energy being transferred from one to the other.

At a time corresponding to Fig. 36-8*c*, the capacitor is fully discharged, and the energy stored in the capacitor is zero. The current in the inductor has reached its maximum value, and all the energy in the circuit is stored in the mag-

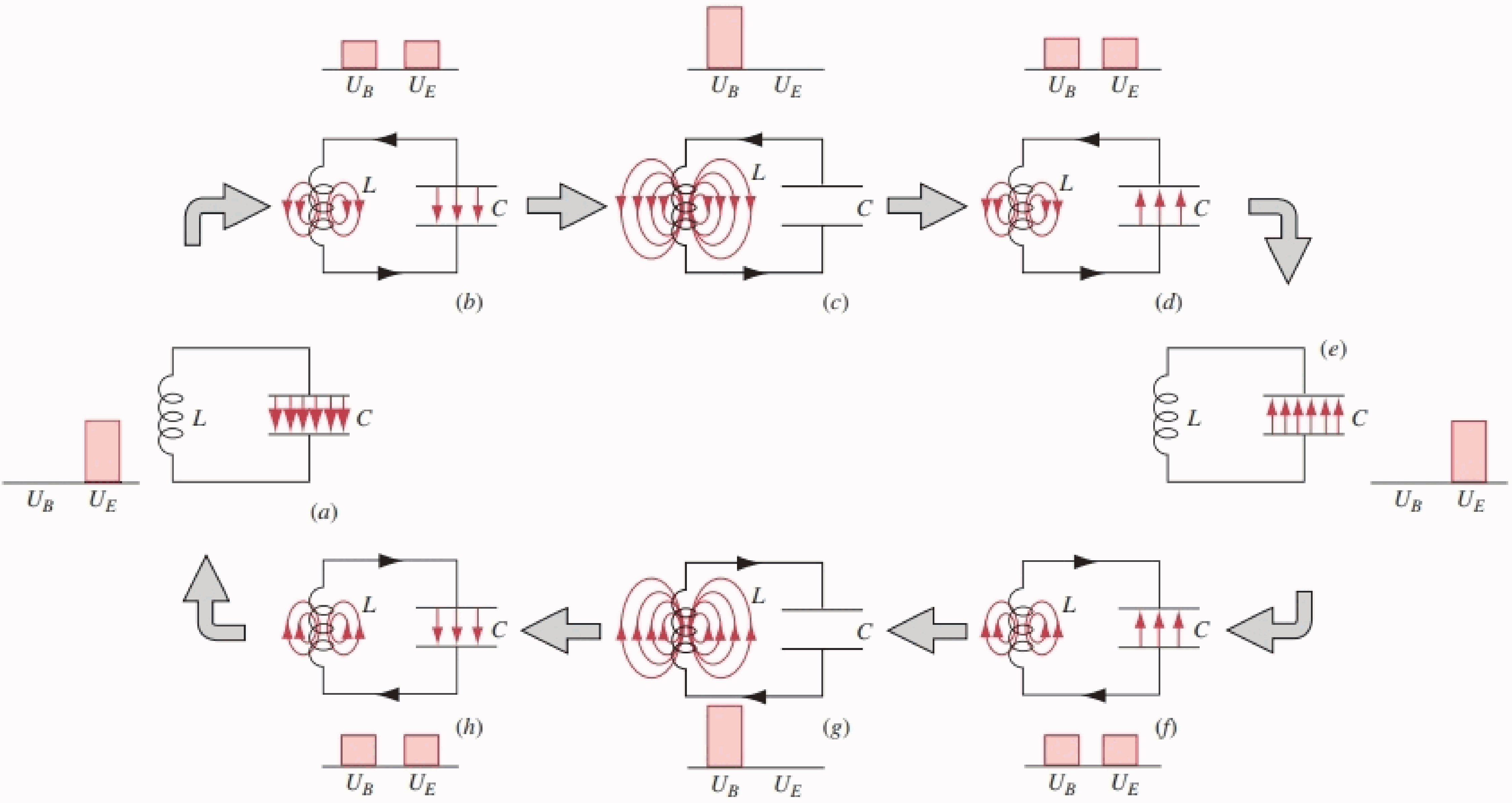


FIGURE 36-8. Eight stages in a single cycle of oscillation of a resistanceless LC circuit. The bar graphs show the stored magnetic and electric energies.

netic field of the inductor. Note that, even though $q = 0$ at this instant, dq/dt differs from zero because charge is flowing.

The current in the inductor continues to transport charge from the top plate of the capacitor to the bottom plate, as in Fig. 36-8*d*; energy is now flowing from the inductor back into the capacitor as its electric field builds up again. Eventually (see Fig. 36-8*e*) all the energy has been transferred back to the capacitor, which is now fully charged but in the opposite sense of Fig. 36-8*a*. The situation continues as the capacitor now discharges until the energy is completely back with the inductor, the magnetic field and the corresponding energy having their maximum values (Fig. 36-8*g*). Finally, the current in the inductor charges the capacitor once again until the capacitor is fully charged and the circuit is back in its original condition (Fig. 36-8*a*). The process then begins again, and the cycle repeats indefinitely. In the absence of resistance, which would cause energy to be dissipated, the charge and current return to their same maximum values in each cycle.

The oscillation of the LC circuit takes place with a definite frequency f (measured in Hz) corresponding to an angular frequency ω ($=2\pi f$ and measured in rad/s). As we discuss in the next section, ω is determined by L and C. By suitable choices of L and C, we can build oscillating circuits with frequencies that range from below audio frequencies (10 Hz) to above microwave frequencies (10 GHz).

To determine the charge q as a function of the time, we can measure the variable potential difference $\Delta V_C(t)$ that exists across the capacitor C, which is related to the charge q by

$$\Delta V_C = \frac{1}{C}q.$$

We can determine the current by inserting into the circuit a resistor R so small that its effect on the circuit is negligible. The potential difference $\Delta V_R(t)$ across R is proportional to the current, according to

$$\Delta V_R = iR.$$

If we were to display $\Delta V_C(t)$ and $\Delta V_R(t)$, such as on the screen of an oscilloscope, the result might be similar to that shown in Fig. 36-9.

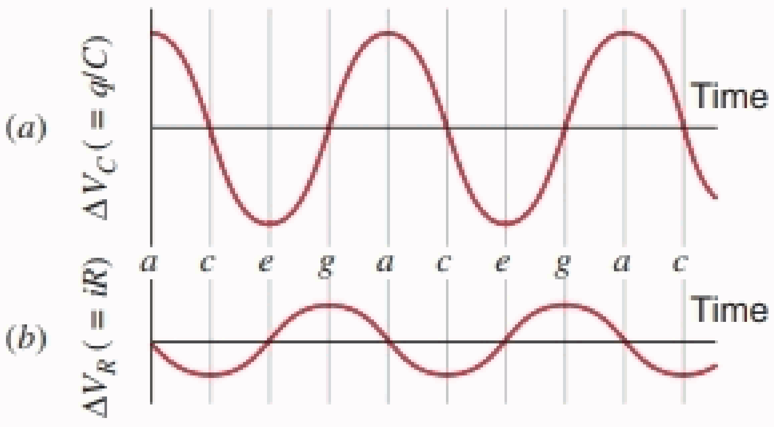


FIGURE 36-9. (*a*) The potential difference across the capacitor in the circuit of Fig. 36-8 as a function of time. This quantity is proportional to the charge on the capacitor. (*b*) The potential difference across a small resistor inserted into the circuit of Fig. 36-8. This quantity is proportional to the current in the circuit. The letters indicate the corresponding stages in the oscillation of Fig. 36-8.

SAMPLE PROBLEM 36-8. A 1.5-μF capacitor is charged to 57 V. The charging battery is then disconnected and a 12-mH coil is connected across the capacitor, so that *LC* oscillations occur. What is the maximum current in the coil? Assume that the circuit contains no resistance.

Solution From the conservation-of-energy principle, the maximum stored energy in the capacitor must equal the maximum stored energy in the inductor (but recall that the two maxima do *not* occur at the same time). Using Eqs. 36-21 and 36-24, we obtain

$$\frac{q_\mathrm{m}^2}{2C} = \tfrac{1}{2}Li_\mathrm{m}^2,$$

where i_m is the maximum current and q_m is the maximum charge. Note that the maximum current and maximum charge do not occur at the same time but are one-fourth of a cycle apart; see Figs. 36-8 and 36-9. Solving for i_m and substituting CV for q_m, we find

$$i_\mathrm{m} = V\sqrt{\frac{C}{L}} = (57\ \mathrm{V})\sqrt{\frac{1.5\times10^{-6}\ \mathrm{F}}{12\times10^{-3}\ \mathrm{H}}} = 0.64\ \mathrm{A}.$$

Analogy to Simple Harmonic Motion

Figure 12-5 shows that in an oscillating block–spring system, as in an oscillating *LC* circuit, two kinds of energy occur. One is potential energy of the compressed or extended spring; the other is kinetic energy of the moving block. These are given by the familiar formulas in the first column of Table 36-1. The table suggests that a capacitor is in some way like a spring, an inductor is like a massive object (the block), and certain electromagnetic quantities "correspond" to certain mechanical ones—namely,

$$\begin{aligned} &q \text{ corresponds to } x, && i \text{ corresponds to } v, \\ &1/C \text{ corresponds to } k, && L \text{ corresponds to } m. \end{aligned} \tag{36-25}$$

Comparison of Fig. 36-8, which shows the oscillations of a resistanceless *LC* circuit, with Fig. 12-5, which shows the oscillations in a frictionless block–spring system, indicates how close the correspondence is. Note how v and i correspond in the two figures, as do x and q. Note also how in each case the energy alternates between two forms, magnetic and electric for the *LC* system, and kinetic and potential for the block–spring system.

In Section 17-3 we saw that the natural angular frequency of a mechanical simple harmonic oscillator is

$$\omega = 2\pi f = \sqrt{\frac{k}{m}}.$$

TABLE 36-1 **Energy in Oscillating Systems**

Mechanical		*Electromagnetic*	
Spring	$U_s = \frac{1}{2}kx^2$	Capacitor	$U_E = \frac{1}{2}C^{-1}q^2$
Block	$K = \frac{1}{2}mv^2$	Inductor	$U_B = \frac{1}{2}Li^2$
	$v = dx/dt$		$i = dq/dt$

The correspondence between the two systems suggests that to find the frequency of oscillation of a (resistanceless) *LC* circuit, k should be replaced by $1/C$ and m by L, which gives

$$\omega = 2\pi f = \sqrt{\frac{1}{LC}}. \tag{36-26}$$

This formula can also be derived from a rigorous analysis of the electromagnetic oscillation, as shown in the next section.

36-6 ELECTROMAGNETIC OSCILLATIONS: QUANTITATIVE

We now derive an expression for the frequency of oscillation of a (resistanceless) *LC* circuit using the conservation-of-energy principle. The total energy U present at any instant in an oscillating *LC* circuit is

$$U = U_B + U_E = \frac{1}{2}Li^2 + \frac{1}{2}\frac{q^2}{C}, \tag{36-27}$$

which indicates that at any arbitrary time the energy is stored partly in the magnetic field of the inductor and partly in the electric field of the capacitor. If we assume the circuit resistance to be zero, no energy is dissipated and U remains constant with time, even though i and q vary. In more formal language, dU/dt must be zero. This leads to

$$\frac{dU}{dt} = \frac{d}{dt}\left(\frac{1}{2}Li^2 + \frac{1}{2}\frac{q^2}{C}\right) = Li\frac{di}{dt} + \frac{q}{C}\frac{dq}{dt} = 0. \tag{36-28}$$

We let q represent the charge on a particular plate of the capacitor (for instance, the upper plate in Fig. 36-8), and i then represents the rate at which charge flows into that plate (so that $i > 0$ when positive charge flows into the plate). In this case

$$i = \frac{dq}{dt} \quad \text{and} \quad \frac{di}{dt} = \frac{d^2q}{dt^2},$$

and substituting into Eq. 36-28 we obtain

$$\frac{d^2q}{dt^2} + \frac{1}{LC}q = 0. \tag{36-29}$$

Equation 36-29 describes the oscillations of a (resistanceless) *LC* circuit. To solve it, note the similarity of Eq. 17-4,

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0, \tag{36-30}$$

which describes the mechanical oscillation of a particle on a spring. Fundamentally, it is by comparing these two equations that the correspondences of Eq. 36-25 arise.

The solution of Eq. 36-30 obtained in Chapter 17 was

$$x = x_\mathrm{m}\cos(\omega t + \phi),$$

where x_m is the amplitude of the motion and ϕ is an arbitrary phase constant. Since q corresponds to x, we can write the solution of Eq. 36-29 as

$$q = q_m \cos(\omega t + \phi), \tag{36-31}$$

where ω is the still unknown angular frequency of the electromagnetic oscillations.

We can test whether Eq. 36-31 is indeed a solution of Eq. 36-29 by substituting it and its second derivative in that equation. To find the second derivative, we write

$$\frac{dq}{dt} = i = -\omega q_m \sin(\omega t + \phi) \tag{36-32}$$

and

$$\frac{d^2q}{dt^2} = -\omega^2 q_m \cos(\omega t + \phi). \tag{36-33}$$

Substituting q and d^2q/dt^2 into Eq. 36-29 yields

$$-\omega^2 q_m \cos(\omega t + \phi) + \frac{1}{LC} q_m \cos(\omega t + \phi) = 0.$$

Canceling $q_m \cos(\omega t + \phi)$ and rearranging leads to

$$\omega = \sqrt{\frac{1}{LC}}. \tag{36-34}$$

Thus, if ω is given the value $1/\sqrt{LC}$, Eq. 36-31 is indeed a solution of Eq. 36-29. This expression for ω agrees with Eq. 36-26, which we arrived at by the correspondence between mechanical and electromagnetic oscillations.

The phase constant ϕ in Eq. 36-31 is determined by the conditions at $t = 0$. If the initial condition is as represented by Fig. 36-8*a*, then we put $\phi = 0$ in order that Eq. 36-31 may predict $q = q_m$ at $t = 0$. What initial physical condition is implied by $\phi = 90°$? $180°$? $270°$? Which of the states shown in Fig. 36-8 correspond to these choices of ϕ?

The stored electric energy in the LC circuit, using Eq. 36-31, is

$$U_E = \frac{1}{2}\frac{q^2}{C} = \frac{q_m^2}{2C}\cos^2(\omega t + \phi), \tag{36-35}$$

and the magnetic energy, using Eq. 36-32, is

$$U_B = \tfrac{1}{2}Li^2 = \tfrac{1}{2}L\omega^2 q_m^2 \sin^2(\omega t + \phi).$$

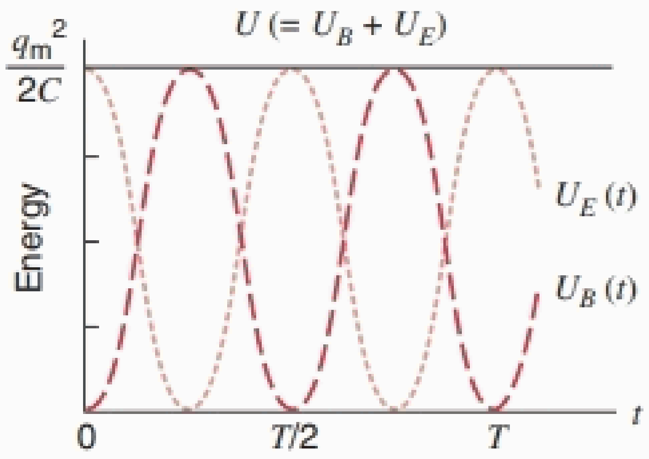


FIGURE 36-10. The stored magnetic energy and electric energy and their sum in an LC circuit as functions of the time. T $(= 2\pi/\omega)$ is the period of the oscillation.

Substituting Eq. 36-34 for ω into this last equation yields

$$U_B = \frac{q_m^2}{2C}\sin^2(\omega t + \phi). \tag{36-36}$$

Figure 36-10 shows plots of $U_E(t)$ and $U_B(t)$ for the case of $\phi = 0$. Note that (1) the maximum values of U_E and U_B are the same $(= q_m^2/2C)$; (2) the sum of U_E and U_B is a constant $(= q_m^2/2C)$; (3) when U_E has its maximum value, U_B is zero and conversely; and (4) U_B and U_E each reach their maximum value *twice* during each cycle. This analysis supports the qualitative analysis of Section 36-5. Compare this discussion with that given in Section 17-4 for the energy transfers in a mechanical simple harmonic oscillator.

SAMPLE PROBLEM 36-9. (*a*) In an oscillating LC circuit, what value of charge, expressed in terms of the maximum charge, is present on the capacitor when the energy is shared equally between the electric and the magnetic field? (*b*) At what time t will this condition occur, assuming the capacitor to be fully charged initially? Assume that $L = 12$ mH and $C = 1.7\ \mu$F.

Solution (*a*) The stored energy U_E and the *maximum* stored energy U_m in the capacitor are, respectively,

$$U_E = \frac{q^2}{2C} \quad \text{and} \quad U_m = \frac{q_m^2}{2C}.$$

Substituting $U_E = \frac{1}{2}U_m$ yields

$$\frac{q^2}{2C} = \frac{1}{2}\frac{q_m^2}{2C}$$

or

$$q = \frac{q_m}{\sqrt{2}}.$$

(*b*) Since $\phi = 0$ in Eq. 36-31 because $q = q_m$ at $t = 0$, we have

$$q = q_m \cos\omega t = \frac{q_m}{\sqrt{2}},$$

which leads to

$$\omega t = \cos^{-1}\frac{1}{\sqrt{2}} = \frac{\pi}{4}$$

or, using $\omega = 1/\sqrt{LC}$,

$$t = \frac{\pi}{4\omega} = \frac{\pi\sqrt{LC}}{4} = \frac{\pi\sqrt{(12 \times 10^{-3}\ \text{H})(1.7 \times 10^{-6}\ \text{F})}}{4}$$
$$= 1.1 \times 10^{-4}\ \text{s} = 110\ \mu\text{s}.$$

36-7 DAMPED AND FORCED OSCILLATIONS

A resistance R is always present in any real LC circuit. When we take this resistance into account, we find that the total electromagnetic energy U is not constant but decreases with time as it is dissipated as internal energy in the resistor. As we shall see, the analogy with the damped block–spring oscillator of Section 17-7 is exact. As before, we have

$$U = U_B + U_E = \frac{1}{2}Li^2 + \frac{q^2}{2C}. \qquad (36\text{-}37)$$

U is no longer constant but rather

$$\frac{dU}{dt} = -i^2R, \qquad (36\text{-}38)$$

the negative sign signifying that the stored energy U decreases with time, being converted to internal energy in the resistor at the rate i^2R. Differentiating Eq. 36-37 and combining the result with Eq. 36-38, we have

$$-i^2R = Li\frac{di}{dt} + \frac{q}{C}\frac{dq}{dt}.$$

Substituting dq/dt for i and d^2q/dt^2 for di/dt and dividing by i, we obtain, after rearranging terms,

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = 0, \qquad (36\text{-}39)$$

which describes the damped LC oscillations. If we put $R = 0$, Eq. 36-39 reduces, as it must, to Eq. 36-29, which describes the undamped LC oscillations.

We state without proof that the general solution of Eq. 38-39 can be written in the form

$$q = q_m e^{-Rt/2L}\cos(\omega' t + \phi), \qquad (36\text{-}40)$$

in which

$$\omega' = \sqrt{\omega^2 - (R/2L)^2}. \qquad (36\text{-}41)$$

Using the analogies of Eq. 36-25, we see that Eq. 36-40 is the exact equivalent of Eq. 17-39, the equation for the displacement as a function of time in damped simple harmonic motion. Comparing Eq. 36-41 with Eq. 17-40, we see that the resistance R corresponds to the damping constant b of the damped mechanical oscillator.

Figure 36-11 shows the current in a damped LC circuit as a function of the time. (Compare Fig. 17-16*b*.) The current oscillates sinusoidally with frequency ω', and the current amplitude decreases exponentially with time. The frequency ω' is strictly less than the frequency ($\omega = 1/\sqrt{LC}$) of the undamped oscillations, but for most cases of interest we can put $\omega' = \omega$ with negligible error.

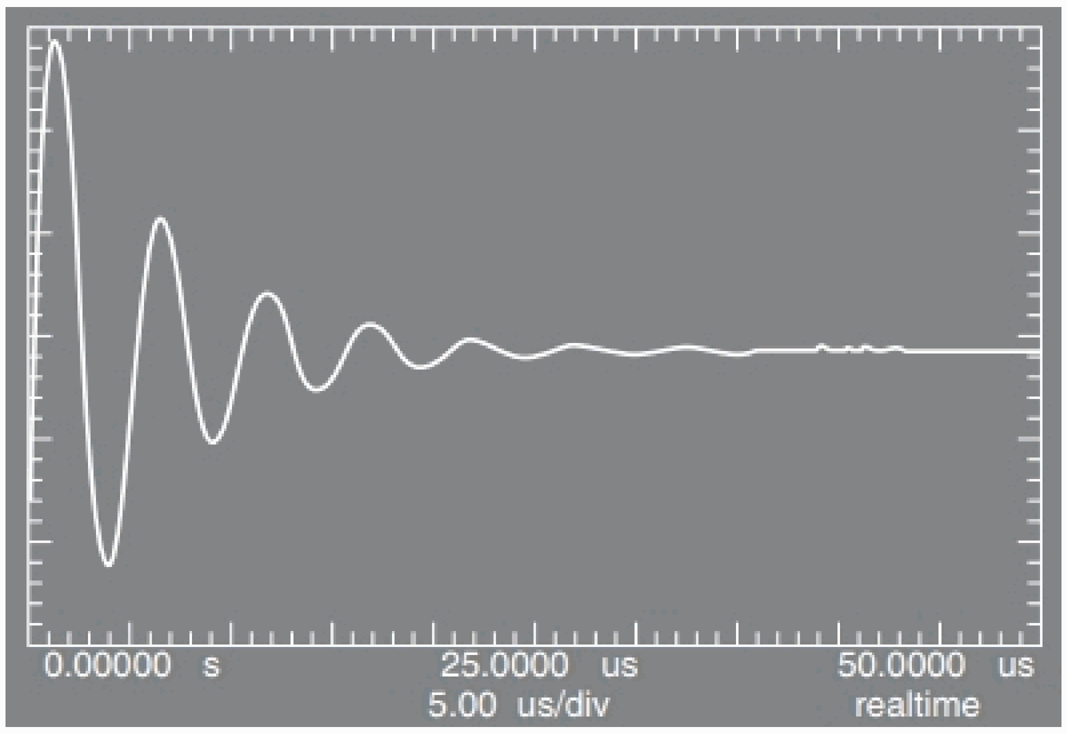


FIGURE 36-11. An oscilloscope trace showing the oscillation of a damped LC circuit. The oscillation decreases in amplitude because energy is dissipated in the resistance of the circuit.

SAMPLE PROBLEM 36-10. A circuit has $L = 12$ mH, $C = 1.6\ \mu$F, and $R = 1.5\ \Omega$. (*a*) After what time t will the amplitude of the charge oscillations drop to one-half of its initial value? (*b*) To how many periods of oscillation does this correspond?

Solution (*a*) This will occur when the amplitude factor $e^{-Rt/2L}$ in Eq. 36-40 has the value 1/2, or

$$e^{-Rt/2L} = \tfrac{1}{2}.$$

Taking the natural logarithm of each side gives

$$-Rt/2L = \ln\tfrac{1}{2} = -\ln 2,$$

or, solving for t,

$$t = \frac{2L}{R}\ln 2 = \frac{(2)(12\times 10^{-3}\ \text{H})}{1.5\ \Omega}\ln 2 = 0.011\ \text{s}.$$

(*b*) The number of oscillations is the elapsed time divided by the period, which is related to the angular frequency ω by $T = 2\pi/\omega$. The angular frequency is

$$\omega = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{(12\times 10^{-3}\ \text{H})(1.6\times 10^{-6}\ \text{F})}} = 7200\ \text{rad/s}.$$

The period of oscillation is then

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{7200\ \text{rad/s}} = 8.7\times 10^{-4}\ \text{s}.$$

The elapsed time, expressed in terms of the period of oscillation, is then

$$\frac{t}{T} = \frac{0.011\ \text{s}}{8.7\times 10^{-4}\ \text{s}} \approx 13.$$

The amplitude drops to one-half after about 13 cycles of oscillation. By comparison, the damping in this example is less severe than that shown in Fig. 36-11, where the amplitude drops to one-half in about one cycle.

In the sample problem, we have used ω rather than ω'. From Eq. 36-41, we calculate $\omega - \omega' = 0.27$ rad/s, and so we make a negligible error in using ω.

Forced Oscillations and Resonance

Consider a damped LC circuit containing a resistance R. If the damping is small, the circuit oscillates at the frequency $\omega = 1/\sqrt{LC}$, which we call the *natural frequency* of the system.

Suppose now that we drive the circuit with a time-varying emf given by

$$\mathcal{E} = \mathcal{E}_m \cos\omega'' t, \qquad (36\text{-}42)$$

using an external generator. Here ω'', which can be varied at will, is the frequency of this external source. We describe such oscillations as *forced.* When the emf described by Eq. 36-42 is first applied, time-varying transient currents appear

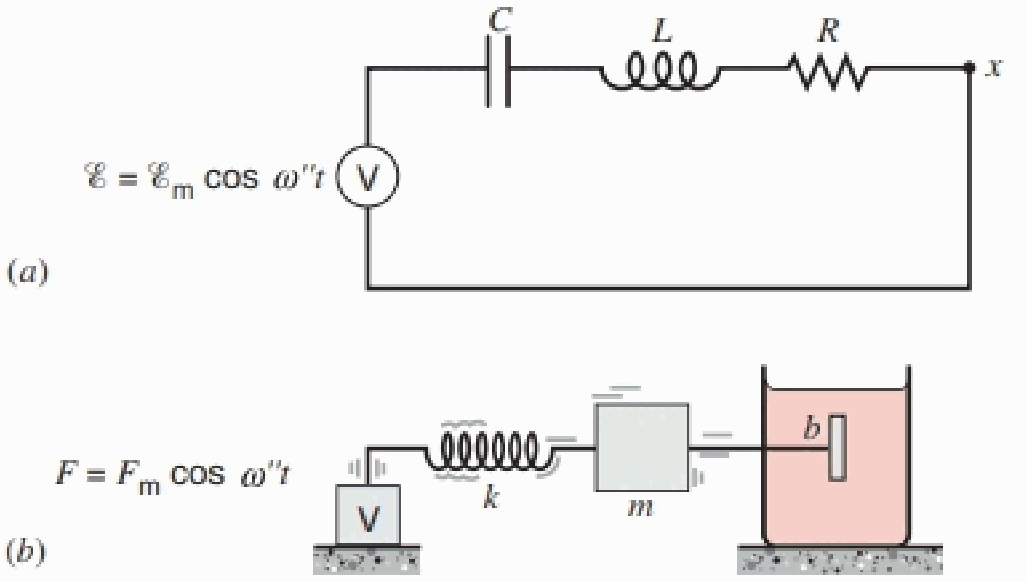


Figure 36-12. (*a*) Electromagnetic oscillations of a circuit are driven at an angular frequency ω''. (*b*) Mechanical oscillations of a spring system are driven at an angular frequency ω''. Corresponding elements of the two systems are drawn opposite each other.

in the circuit. Our interest, however, is in the sinusoidal currents that exist in the circuit after these initial transients have died away. Whatever the natural frequency ω may be, *these oscillations of charge, current, or potential difference in the circuit must occur at the external driving frequency* ω''.

Figure 36-12 compares the electromagnetic oscillating system with a corresponding mechanical system. A vibrator V, which imposes an external alternating force, corresponds to generator V, which imposes an external alternating emf. Other quantities "correspond" as before (see Table 36-1): displacement to charge and velocity to current. The inductance L, which opposes changes in current, corresponds to the mass (inertia) m, which opposes changes in velocity. The spring constant k and the inverse capacitance C^{-1} represent the "stiffness" of their systems, giving, respectively, the response (displacement) of the spring to the force and the response (charge) of the capacitor to the emf.

In Chapter 37 we derive the solution for the current in the circuit of Fig. 36-12*a*, which we can write in the form

$$i = i_m \sin(\omega'' t - \phi). \qquad (36\text{-}43)$$

The current amplitude i_m in Eq. 36-43 is a measure of the response of the circuit of Fig. 36-12*a* to the driving emf. It is reasonable to suppose (from experience in pushing swings, for example) that i_m is large when the driving frequency ω'' is close to the natural frequency ω of the system. In other words, we expect that a plot of i_m versus ω'' exhibits a maximum when

$$\omega'' = \omega = 1/\sqrt{LC}, \qquad (36\text{-}44)$$

which we call the *resonance* condition.

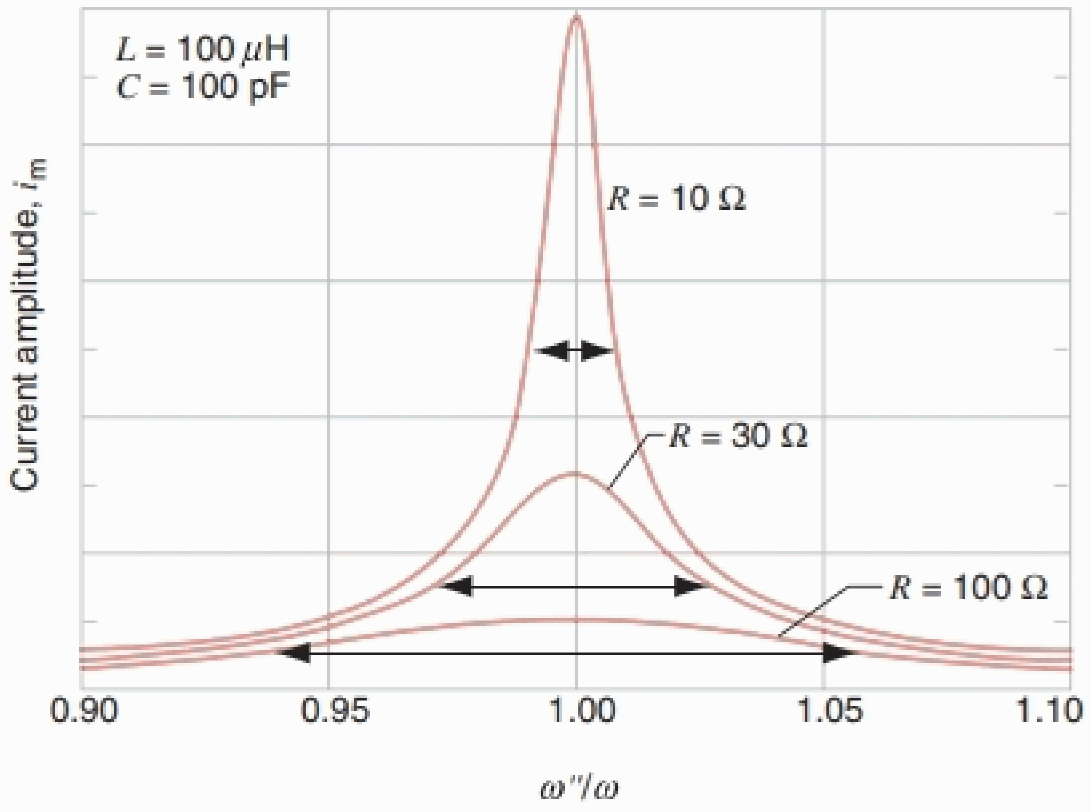


Figure 36-13. Resonance curves for the forced oscillating circuit of Fig. 36-12*a*. The three curves correspond to different values of the resistance of the circuit. The horizontal arrows indicate the width or "sharpness" of each resonance.

Figure 36-13 shows three plots of i_m as a function of the ratio ω''/ω, each plot corresponding to a different value of the resistance R. We see that each of these peaks does indeed have a maximum value when the resonance condition of Eq. 36-44 is satisfied. Note that as R is decreased, the resonance peak becomes sharper, as shown by the three horizontal arrows drawn at the half-maximum level of each curve.

Figure 36-13 suggests the common experience of tuning a radio set. In turning the tuning knob, we are adjusting the natural frequency ω of an internal LC circuit to match the driving frequency ω'' of the signal transmitted by the antenna of the broadcasting station; we are looking for resonance. In a metropolitan area, where there are many signals whose frequencies are often close together, sharpness of tuning becomes important.

Figure 36-13 is similar to Fig. 17-19, which shows resonance peaks for the forced oscillations of a mechanical oscillator such as that of Fig. 36-12*b*. In this case also, the maximum response occurs when $\omega'' = \omega$, and the resonance peaks become sharper as the damping factor (the coefficient b) is reduced. Note that the curves of Fig. 36-13 and of Fig. 17-19 are not exactly alike. The former is a plot of current amplitude, whereas the latter is a plot of displacement amplitude. The mechanical variable that corresponds to current is not displacement but velocity. Nevertheless, both sets of curves illustrate the resonance phenomenon.

MULTIPLE CHOICE

36-1 Inductance

36-2 Calculating the Inductance

1. Two identical inductors of inductance L are connected in series.

(a) If the inductors are *very* far apart, then the effective inductance of the arrangement will be

(A) $L/2$. (B) $L/\sqrt{2}$. (C) $\sqrt{2}L$. (D) $2L$.

(b) If the two inductors are moved closer together, the effective inductance will

(A) increase. (B) decrease.
(C) remain the same.
(D) change, but the answer depends on the orientation of the inductors.

2. Two identical inductors of inductance L are connected in parallel.

(a) If the inductors are *very* far apart, then the effective inductance of the arrangement will be

(A) $L/2$. (B) $L/\sqrt{2}$. (C) $\sqrt{2}L$. (D) $2L$.

(b) If the two inductors are moved closer together, the effective inductance will

(A) increase. (B) decrease.
(C) remain the same.
(D) change, but the answer depends on the orientation of the inductors.

3. A single wire loop has an inductance L_0. If a second wire loop is added on top of the first, the new inductance will be

(A) $4L_0$. (B) $2L_0$. (C) $\sqrt{2}L_0$. (D) L_0^2.

4. How does the inductance per unit length for a real solenoid near its center compare to the inductance per unit length near its ends?

(A) The inductance per unit length is larger near the center.
(B) The inductance per unit length is larger near the ends.
(C) The inductance per unit length is the same everywhere.

5. An inductor can be made out of a metal "Slinky," which is basically a flexible solenoid. If the "Slinky" is stretched out to twice the original length, then the inductance will

(A) change to $L_0/2$. (B) change to $L_0/4$.
(C) change to $\sqrt{L_0}$. (D) remain the same.

6. An inductor has an inductance L_0. A second inductor is identical to the first except that it is twice as large; it is magnified by a factor of two. What is the ratio of the inductance of the magnified inductor to the original inductor?

(A) 4 (B) 2 (C) 1 (D) 1/2
(E) The answer depends on the geometry of the inductor.

36-3 *LR* Circuits

7. A resistor, inductor, switch, and battery are in series. Originally the switch is open. The switch is then closed, allowing current to flow.

(a) Before the system reaches a steady state, the potential difference across the resistor is

(A) increasing. (B) decreasing. (C) fixed.

(b) Before the system reaches a steady state, the potential difference across the inductor is

(A) increasing. (B) decreasing. (C) fixed.

8. An ideal inductor is connected in series with a resistor and an ideal battery. The battery supplies energy at a rate $P(t)$, the resistor dissipates energy at a rate $P_R(t)$, and the inductor stores energy at a rate $P_L(t)$. What can be concluded about the relationship between $P_R(t)$ and $P_L(t)$?

(A) $P_R(t) > P_L(t)$ for all times t while charging.
(B) $P_R(t) = P_L(t)$ for all times t while charging.
(C) $P_R(t) < P_L(t)$ for all times t while charging.
(D) $P_R(t) > P_L(t)$ only near the beginning of the charging.
(E) $P_R(t) < P_L(t)$ only near the beginning of the charging.

36-4 Energy Storage in a Magnetic Field

9. Consider again the two inductors of Multiple-choice question 6, and assume they carry the same current.

(a) What is the ratio of the magnetic energy density of the magnified inductor to that of the original inductor?

(A) 1/4 (B) 1/2 (C) 2 (D) 4

(b) What is the ratio of the total stored energy of the magnified inductor to that of the original inductor?

(A) 1/4 (B) 1/2 (C) 2 (D) 4

36-5 Electromagnetic Oscillations: Qualitative

10. A simple LC circuit consists of an inductor L and a capacitor C in series. The circuit oscillates at a frequency f_0. A second identical inductor and a second identical capacitor are added in series to the circuit; assume the inductors are far enough apart so that their magnetic fields do not interfere. What is the new frequency of oscillation for this system?

(A) $4f_0$ (B) $2f_0$ (C) f_0
(D) $f_0/2$ (E) $f_0/4$

11. The frequency of an LC oscillator is f_0. The plates of the parallel plate capacitor are then pulled apart to twice the original distance. What is the new frequency of oscillation?

(A) $2f_0$ (B) $\sqrt{2}f_0$ (C) $f_0/\sqrt{2}$ (D) $f_0/2$

12. Which of the following changes will increase the frequency of an LC oscillator? (*There may be more than one correct answer!*)

(A) Inserting a dielectric slab into the capacitor
(B) Inserting a paramagnetic substance into the inductor
(C) Inserting a ferromagnetic substance into the inductor
(D) Inserting a diamagnetic substance into the inductor

36-6 Electromagnetic Oscillations: Quantitative

13. An LC circuit oscillates with frequency f. When the current through the inductor is $i = i_{\max}/2$, the charge on the capacitor is

(A) $q = q_{\max}$. (B) $q = \sqrt{3}q_{\max}/2$.
(C) $q = \sqrt{2}q_{\max}/2$. (D) $q = q_{\max}/2$.

14. An LC circuit originally oscillates with frequency f and maximum current $i_{\max}$. If the total energy present in the circuit is doubled, then

(a) the period of oscillation will increase by a factor of

(A) 4. (B) $\sqrt{8}$. (C) 2. (D) $\sqrt{2}$.
(E) 1 (it remains unchanged).

(b) the maximum charge that will build up on the capacitor will increase by a factor of

(A) 4. (B) $\sqrt{8}$. (C) 2. (D) $\sqrt{2}$.
(E) 1 (it remains unchanged).

(*c*) the maximum current that will occur in the circuit will increase by a factor of
(A) 4. (B) $\sqrt{8}$. (C) 2. (D) $\sqrt{2}$.
(E) 1 (it remains unchanged).

36-7 Damped and Forced Oscillations

15. When does the energy of an *LRC* circuit decrease at the fastest rate?

(A) When the charge on the capacitor is maximum
(B) When the charge on the capacitor is minimum
(C) While the induced emf in the inductor is a maximum
(D) When Joule heating in the resistor is a minimum

QUESTIONS

1. Show that the dimensions of the two expressions for L, $N\Phi_B/i$ (Eq. 36-4) and $\mathcal{E}_L/(di/dt)$ (Eq. 36-1), are the same.

2. If the flux passing through each turn of a coil is the same, the inductance of the coil may be calculated from $L = N\Phi_B/i$ (Eq. 36-4). How might one compute L for a coil for which this assumption is not valid?

3. Give examples of how the flux linked by a coil can change due to stretching or compression of the coil.

4. You want to wind a coil so that it has resistance but essentially no inductance. How would you do it?

5. A long cylinder is wound from left to right with one layer of wire, giving it n turns per unit length with an inductance of L_1, as in Fig. 36-14*a*. If the winding is now continued, in the same sense but returning from right to left as in Figure 36-14*b*, so as to give a second layer also of n turns per unit length, then what is the value of the inductance? Explain.

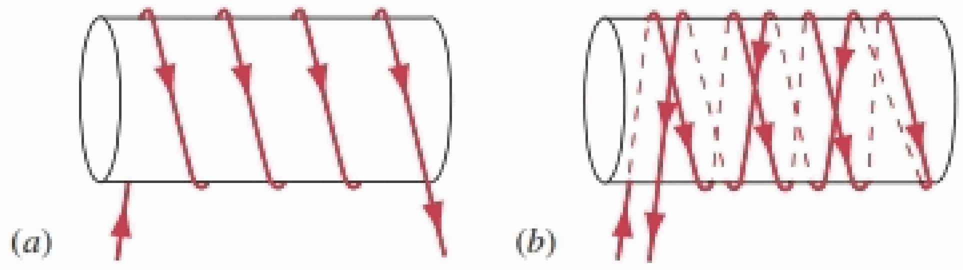


FIGURE 36-14. Question 5.

6. Explain why the inductance of a coaxial cable is expected to increase when the radius of the outer conductor is increased, with the radius of the inner conductor remaining fixed.

7. You are given a length l of copper wire. How would you arrange it to obtain the maximum inductance?

8. Explain how a long, straight wire can show induction effects. How would you go about looking for them?

9. A steady current is set up in a coil with a very large inductive time constant. When the current is interrupted with a switch, a heavy arc tends to appear at the switch blades. Explain why. (Note: Interrupting currents in highly inductive circuits can be destructive and dangerous.)

10. Suppose that you connect an ideal (that is, essentially resistanceless) coil across an ideal (again, essentially resistanceless) battery. You might think that, because there is no resistance in the circuit, the current would jump at once to a very large value. On the other hand, you might think that, because the inductive time constant ($=L/R$) is very large, the current would rise very slowly, if at all. What actually happens?

11. In an *LR* circuit like that of Fig. 36-4, can the induced emf ever be larger than the battery emf?

12. In an *LR* circuit like that of Fig. 36-4, is the current in the resistor always the same as the current in the inductor?

13. In the circuit of Fig. 36-4, the induced emf is a maximum at the instant the switch is closed on a. How can this be, since there is no current in the inductor at this instant?

14. Does the time required for the current in a particular *LR* circuit to build up to a given fraction of its equilibrium value depend on the value of the applied constant emf?

15. If the current in a source of emf is in the direction of the emf, the energy of the source decreases; if a current is in a direction opposite to the emf (as in charging a battery), the energy of the source increases. Do these statements apply to the inductor in Fig. 36-2?

16. Can the emf in an inductor be in the same sense as the emf of the source, which gives the inductor its magnetic energy?

17. The switch in Fig. 36-4, having been closed on a for a "long" time, is thrown to b. What happens to the energy that is stored in the inductor?

18. A coil has a (measured) inductance L and a (measured) resistance R. Is its inductive time constant necessarily given by $\tau_L = L/R$? Bear in mind that we derived that equation (see Fig. 36-4) for a situation in which the inductive and resistive elements are physically separated. Discuss.

19. Figure 36-5*a* and Fig. 31-21*b* are plots of $\Delta V_R(t)$ for, respectively, an *LR* circuit and an *RC* circuit. Why are these two curves so different? Account for each in terms of physical processes going on in the appropriate circuit.

20. Two solenoids, *A* and *B*, have the same diameter and length and contain only one layer of copper windings, with adjacent turns touching and insulation thickness being negligible. Solenoid *A* contains many turns of fine wire and solenoid *B* contains fewer turns of heavier wire. (*a*) Which solenoid has the larger inductance? (*b*) Which solenoid has the larger inductive time constant? Justify your answers.

21. Can you make an argument based on the manipulation of bar magnets to suggest that energy may be stored in a magnetic field?

22. Draw all the formal analogies you can think of between a parallel-plate capacitor (for electric fields) and a long solenoid (for magnetic fields).

23. In each of the following operations energy is expended. Some of this energy is returnable (can be reconverted) into electrical energy that can be made to do useful work and some becomes unavailable for useful work or is wasted in other ways. In which case will there be the least fraction of returnable electrical energy: (*a*) charging a capacitor, (*b*) charging a storage battery, (*c*) sending a current through a resistor, (*d*) setting up a magnetic field, and (*e*) moving a conductor in a magnetic field?

24. The current in a solenoid is reversed. What changes does this make in the magnetic field $\vec{\mathbf{B}}$ and the energy density u_B at various points along the solenoid axis?

25. Commercial devices such as motors and generators that are involved in the transformation of energy between electrical and mechanical forms involve magnetic rather than electrostatic fields. Why should this be so?

26. Why doesn't the *LC* circuit of Fig. 36-8 simply stop oscillating when the capacitor has been completely discharged?

27. How might you start an *LC* circuit into oscillation with its initial condition being represented by Fig. 36-8*c*? Devise a switching scheme to bring this about.

28. The lower curve *b* in Fig. 36-9 is proportional to the derivative of the upper curve *a*. Explain why.

29. In an oscillating *LC* circuit, assumed resistanceless, what determines (*a*) the frequency and (*b*) the amplitude of the oscillations?

30. In connection with Figs. 36-8*c* and 36-8*g*, explain how there can be a current in the inductor even though there is no charge on the capacitor.

31. In Fig. 36-8, what changes are required if the oscillations are to proceed counterclockwise around the figure?

32. In Fig. 36-8, what phase constants ϕ in Eq. 36-31 would permit the eight circuit situations shown to serve in turn as initial conditions?

33. What constructional difficulties would you encounter if you tried to build an *LC* circuit of the type shown in Fig. 36-8 to oscillate (*a*) at 0.01 Hz or (*b*) at 10 GHz?

34. Two inductors L_1 and L_2 and two capacitors C_1 and C_2 can be connected in series according to the arrangement in Fig. 36-15*a* or Fig. 36-15*b*. Are the frequencies of the two oscillating circuits equal? Consider the two cases (*a*) $C_1 = C_2$, $L_1 = L_2$ and (*b*) $C_1 \neq C_2$, $L_1 \neq L_2$.

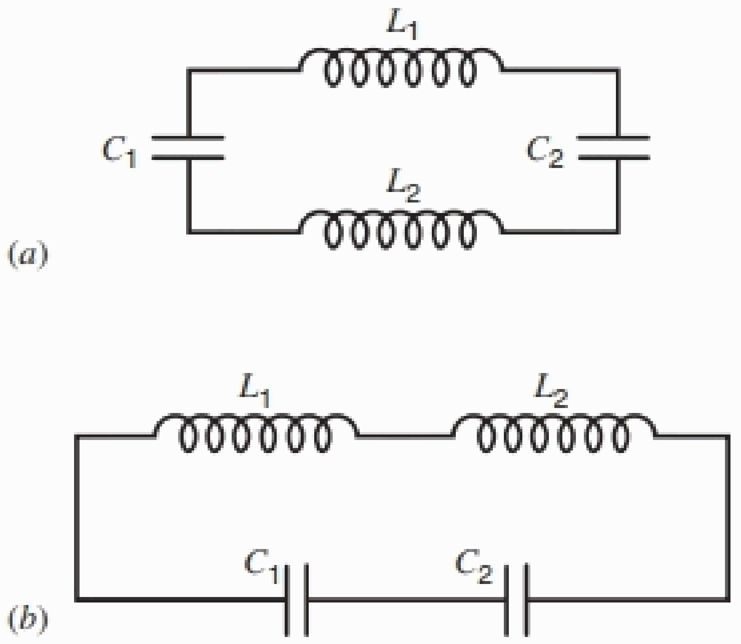


FIGURE 36-15. Question 34.

35. In the mechanical analogy to the oscillating *LC* circuit, what mechanical quantity corresponds to the potential difference?

36. In comparing the electromagnetic oscillating system to a mechanical oscillating system, to what mechanical properties are the following electromagnetic properties analogous: capacitance, resistance, charge, electric field energy, magnetic field energy, inductance, and current?

37. Two springs are joined and connected to an object with mass *m*, the arrangement being free to oscillate on a horizontal frictionless surface as in Fig. 36-16. Sketch the electromagnetic analog of this mechanical oscillating system.

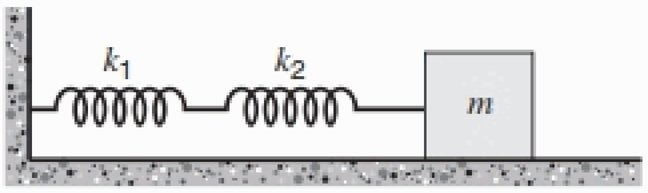


FIGURE 36-16. Question 37.

38. Explain why it is not possible to have (*a*) a real *LC* circuit without resistance, (*b*) a real inductor without inherent capacitance, or (*c*) a real capacitor without inherent inductance. Discuss the practical validity of the *LC* circuit of Fig. 36-8, in which each of the above realities is ignored.

39. All practical *LC* circuits must contain some resistance. However, one can buy a packaged audio oscillator in which the output maintains a constant amplitude indefinitely and does not decay, as it does in Fig. 36-11. How can this happen?

40. What would a resonance curve for $R = 0$ look like if plotted in Fig. 36-13?

41. Can you see any physical reason for assuming that *R* is "small" in Eqs. 36-40 and 36-41? (Hint: Consider what might happen if the damping *R* were so large that Eq. 36-40 would not even go through one cycle of oscillation before *q* was reduced essentially to zero. Could this happen? If so, what do you imagine Fig. 36-11 would look like?)

42. What is the difference between free, damped, and forced oscillating circuits?

43. Tabulate as many mechanical or electrical systems as you can think of that possess a natural frequency, along with the formula for that frequency if given in the text.

44. In an oscillatory radio receiver circuit, is it desirable to have a low or a high *Q*-factor? Explain. (See Problem 15.)

EXERCISES

36-1 Inductance

36-2 Calculating the Inductance

1. The inductance of a close-packed coil of 400 turns is 8.0 mH. Calculate the magnetic flux through the coil when the current is 5.0 mA.

2. A circular coil has a 10.3-cm radius and consists of 34 closely wound turns of wire. An externally produced magnetic field of 2.62 mT is perpendicular to the coil. (*a*) If no current is in the coil, what is the number of flux linkages? (*b*) When the current in the coil is 3.77 A in a certain direction, the net flux through the coil is found to vanish. Find the inductance of the coil.

3. A solenoid is wound with a single layer of insulated copper wire (diameter, 2.52 mm). It is 4.10 cm in diameter and 2.0 m long. What is the inductance per meter for the solenoid near its center? Assume that adjacent wires touch and that insulation thickness is negligible.

4. At a given instant the current and the induced emf in an inductor are as indicated in Fig. 36-17. (*a*) Is the current increasing or decreasing? (*b*) The emf is 17 V, and the rate of change of the current is 25 kA/s; what is the value of the inductance?

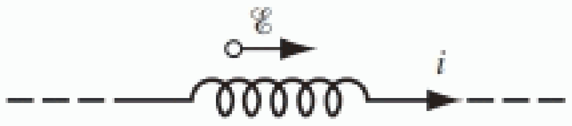


FIGURE 36-17. Exercise 4.

5. The inductance of a closely wound N-turn coil is such that an emf of 3.0 mV is induced when the current changes at the rate 5.0 A/s. A steady current of 8.0 A produces a magnetic flux of 40 μWb through each turn. (*a*) Calculate the inductance of the coil. (*b*) How many turns does the coil have?

6. A toroid having a 5.20-cm square cross section and an inside radius of 15.3 cm has 536 turns of wire and carries a current of 810 mA. Calculate the magnetic flux through a cross section.

7. A solenoid 126 cm long is formed from 1870 windings carrying a current of 4.36 A. The core of the solenoid is filled with iron, and the effective permeability constant is 968. Calculate the inductance of the solenoid, assuming that it can be treated as ideal, with a diameter of 5.45 cm.

8. The current i through a 4.6-H inductor varies with time t as shown on the graph of Fig. 36-18. Calculate the induced emf during the time intervals (*a*) $t = 0$ to $t = 2$ ms, (*b*) $t = 2$ ms to $t = 5$ ms, and (*c*) $t = 5$ ms to $t = 6$ ms. (Ignore the behavior at the ends of the intervals.)

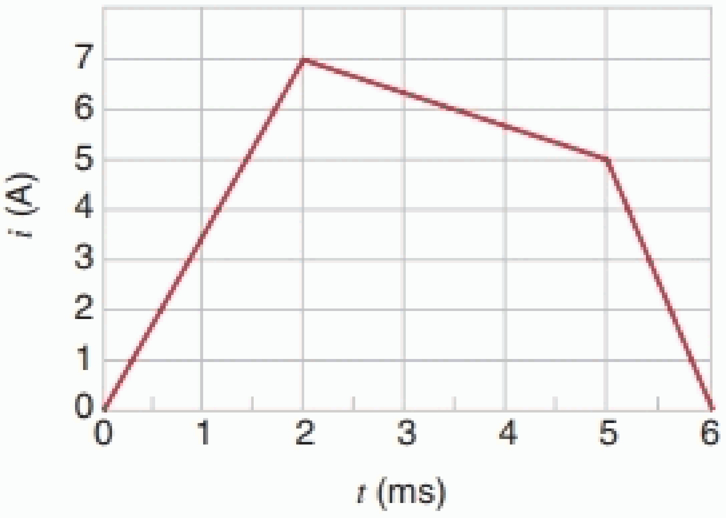


FIGURE 36-18. Exercise 8.

9. Two inductors L_1 and L_2 are connected in parallel and separated by a large distance. (*a*) Show that the equivalent inductance is given from

$$\frac{1}{L_{\text{eq}}} = \frac{1}{L_1} + \frac{1}{L_2}.$$

(*b*) Why must their separation be large for this relationship to hold?

10. Two inductors L_1 and L_2 are connected in series and are separated by a large distance. (*a*) Show that the equivalent inductance is given by

$$L_{\text{eq}} = L_1 + L_2.$$

(*b*) Why must their separation be large for this relationship to hold?

36-3 *LR* Circuits

11. The current in an LR circuit drops from 1.16 A to 10.2 mA in the 1.50 s immediately following removal of the battery from the circuit. If L is 9.44 H, find the resistance R in the circuit.

12. Consider the LR circuit of Fig. 36-4. (*a*) In terms of the battery emf $\mathcal{E}$, what is the induced emf $\mathcal{E}_L$ when the switch has just been closed on a? (*b*) What is $\mathcal{E}_L$ after two time constants? (*c*) After how many time constants will $\mathcal{E}_L$ be just one-half of the battery emf $\mathcal{E}$?

13. The number of flux linkages through a certain coil of 745-mΩ resistance is 26.2 mWb when there is a current of 5.48 A in it. (*a*) Calculate the inductance of the coil. (*b*) If a 6.00-V battery is suddenly connected across the coil, how long will it take for the current to rise from 0 to 2.53 A?

14. (*a*) Show that Eq. 36-12 can be written

$$\frac{di}{i - \mathcal{E}/R} = -\frac{R}{L}\,dt.$$

(*b*) Integrate this equation to obtain Eq. 36-13.

15. Suppose the emf of the battery in the circuit of Fig. 36-4 (with the switch closed on a) varies with time t so the current is given by $i(t) = (3.0\text{ A}) + (5.0\text{ A/s})t$. Take $R = 4.0\ \Omega$, $L = 6.0$ H, and find an expression for the battery emf as a function of time. (Hint: Apply the loop rule.)

16. The current in an LR circuit builds up to one-third of its steady-state value in 5.22 s. Calculate the inductive time constant.

17. A 45-V potential difference is suddenly applied to a coil with $L = 50$ mH and $R = 180\ \Omega$. At what rate is the current increasing after 1.2 ms?

18. At $t = 0$ a battery is connected to an inductor and resistor connected in series. The following table gives the measured potential difference, in volts, across the inductor as a function of time, in ms, following the connection of the battery. Deduce (*a*) the emf of the battery and (*b*) the time constant of the circuit.

t(ms)	ΔV_L(V)	t(ms)	ΔV_L(V)
1.0	18.24	5.0	5.98
2.0	13.8	6.0	4.53
3.0	10.4	7.0	3.43
4.0	7.90	8.0	2.60

19. In Fig. 36-19, $\mathcal{E} = 100$ V, $R_1 = 10\ \Omega$, $R_2 = 20\ \Omega$, $R_3 = 30\ \Omega$, and $L = 2.0$ H. Find the values of i_1 and i_2 (*a*) immediately after switch S is closed, (*b*) a long time later, (*c*) immediately after switch S is opened again, and (*d*) a long time later.

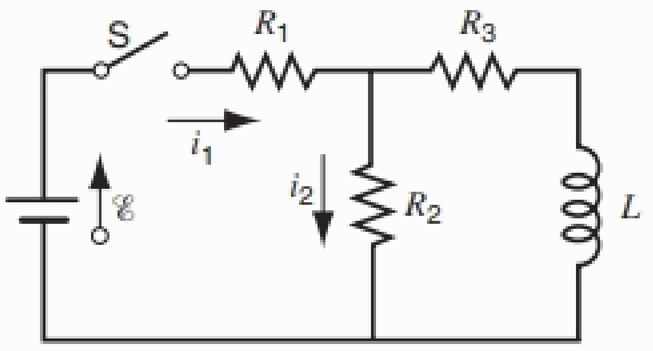


FIGURE 36-19. Exercise 19.

20. A wooden toroidal core with a square cross section has an inner radius of 10 cm and an outer radius of 12 cm. It is wound with one layer of wire (diameter, 0.96 mm; resistance per unit length, 21 mΩ/m). Calculate (*a*) the inductance and (*b*) the inductive time constant. Ignore the thickness of the insulation.

21. In the circuit shown in Fig. 36-20, $\mathcal{E} = 10$ V, $R_1 = 5.0\ \Omega$, $R_2 = 10\ \Omega$, and $L = 5.0$ H. For the two separate conditions (I) switch S just closed and (II) switch S closed for a long time, calculate (*a*) the current i_1 through R_1, (*b*) the current i_2 through R_2, (*c*) the current i through the switch, (*d*) the potential difference across R_2, (*e*) the potential difference across L, and (*f*) di_2/dt.

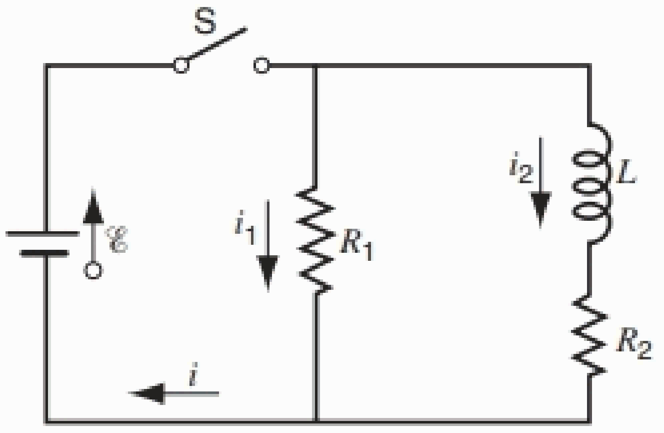


FIGURE 36-20. Exercise 21.

36-4 Energy Storage in a Magnetic Field

22. A 92-mH toroidal inductor encloses a volume of 0.022 m^3. If the average energy density in the toroid is 71 J/m^3, calculate the current.

23. The magnetic energy stored in a certain inductor is 25.3 mJ when the current is 62.0 mA. (*a*) Calculate the inductance. (*b*) What current is required for the magnetic energy to be four times as much?

24. A solenoid 85.3 cm long has a cross-sectional area of 17.2 cm^2. There are 950 turns of wire carrying a current of 6.57 A. (*a*) Calculate the magnetic field energy density inside the solenoid. (*b*) Find the total energy stored in the magnetic field inside the solenoid. (Neglect end effects.)

25. Find the magnetic energy density at the center of a circulating electron in the hydrogen atom (see Sample Problem 33-2).

26. The magnetic field in the interstellar space of our galaxy has a magnitude of about 100 pT. (*a*) Calculate the corresponding energy density, in eV/cm^3. (*b*) How much energy is stored in this field in a cube 10 light-years on edge? (For scale, note that the nearest star, other than the Sun, is 4.3 light-years distant and the "radius" of our galaxy is about 80,000 light-years.)

27. What must be the magnitude of a uniform electric field if it is to have the same energy density as that possessed by a 0.50-T magnetic field?

28. Suppose that the inductive time constant for the circuit of Fig. 36-4 is 37.5 ms and the current in the circuit is zero at time $t = 0$, when the switch is thrown to a. At what time does the rate at which internal energy is increasing in the resistor equal the rate at which energy is being stored in the inductor?

29. A coil is connected in series with a 10.4-kΩ resistor. When a 55.0-V battery is applied to the two, the current reaches a value of 1.96 mA after 5.20 ms. (*a*) Find the inductance of the coil. (*b*) How much energy is stored in the coil at this same moment?

30. For the circuit of Fig. 36-4, assume that $\mathcal{E} = 12.2$ V, $R = 7.34\ \Omega$, and $L = 5.48$ H. The battery is connected at time $t = 0$. (*a*) How much energy is delivered by the battery during the first 2.00 s? (*b*) How much of this energy is stored in the magnetic field of the inductor? (*c*) How much has appeared in the resistor?

31. The magnetic field at the Earth's surface has a strength of about 60 μT. Assuming this to be relatively constant over radial distances that are small compared with the radius of the Earth and neglecting the variations near the magnetic poles, calculate the energy stored in a shell between the Earth's surface and 16 km above the surface.

32. A length of copper wire carries a current of 10 A, uniformly distributed. Calculate (*a*) the magnetic energy density and (*b*) the electric energy density at the surface of the wire. The wire diameter is 2.5 mm and its resistance per unit length is 3.3 Ω/km.

36-5 Electromagnetic Oscillations: Qualitative

33. A 1.48-mH inductor in an *LC* circuit stores a maximum energy of 11.2 μJ. What is the peak current?

34. What is the capacitance of an *LC* circuit if the maximum charge on the capacitor is 1.63 μC and the total energy is 142 μJ?

35. *LC* oscillators have been used in circuits connected to loudspeakers to create some of the sounds of "electronic music." What inductance must be used with a 6.7-μF capacitor to produce a frequency of 10 kHz, near the upper end of the audible range of frequencies?

36. In an oscillating *LC* circuit, $L = 1.13$ mH and $C = 3.88\ \mu$F. The maximum charge on the capacitor is 2.94 μC. Find the maximum current.

37. Consider the circuit shown in Fig. 36-21. With switch S_1 closed and the other two switches open, the circuit has a time constant τ_C. With switch S_2 closed and the other two switches open, the circuit has a time constant τ_L. With switch S_3 closed and the other two switches open, the circuit oscillates with a period T. Show that $T = 2\pi\sqrt{\tau_C \tau_L}$.

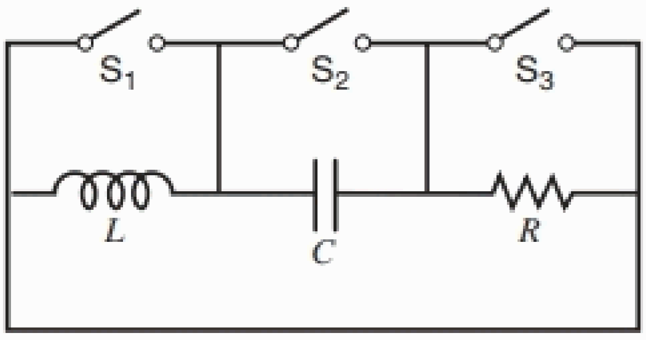


FIGURE 36-21. Exercise 37.

38. You are given a 10.0-mH inductor and two capacitors, of 5.00 μF and 2.00 μF capacitance. List the resonant frequencies that can be generated by connecting these elements in various combinations.

39. A 485-g body oscillates on a spring that, when extended 2.10 mm from equilibrium, has a restoring force of 8.13 N. (*a*) Calculate the angular frequency of oscillation. (*b*) What is its period of oscillation? (*c*) What is the capacitance of the analogous *LC* system if *L* is chosen to be 5.20 H?

36-6 Electromagnetic Oscillations: Quantitative

40. In an *LC* circuit with $L = 52.2$ mH and $C = 4.21$ μF, the current is initially a maximum. How long will it take before the capacitor is fully charged for the first time?

41. For a certain *LC* circuit the total energy is converted from electrical energy in the capacitor to magnetic energy in the inductor in 1.52 μs. (*a*) What is the period of oscillation? (*b*) What is the frequency of oscillation? (*c*) How long after the magnetic energy is a maximum will it be a maximum again?

42. An oscillating *LC* circuit consisting of a 1.13-nF capacitor and a 3.17-mH coil has a peak potential drop of 2.87 V. Find (*a*) the maximum charge on the capacitor, (*b*) the peak current in the circuit, and (*c*) the maximum energy stored in the magnetic field of the coil.

43. An oscillating *LC* circuit is designed to operate at a peak current of 31 mA. The inductance of 42 mH is fixed and the frequency is varied by changing *C*. (*a*) If the capacitor has a maximum peak voltage of 50 V, can the circuit safely operate at a frequency of 1.0 MHz? (*b*) What is the maximum safe operating frequency? (*c*) What is the minimum capacitance?

44. In the circuit shown in Fig. 36-22 the switch has been in position *a* for a long time. It is now thrown to *b*. (*a*) Calculate the frequency of the resulting oscillating current. (*b*) What will be the amplitude of the current oscillations?

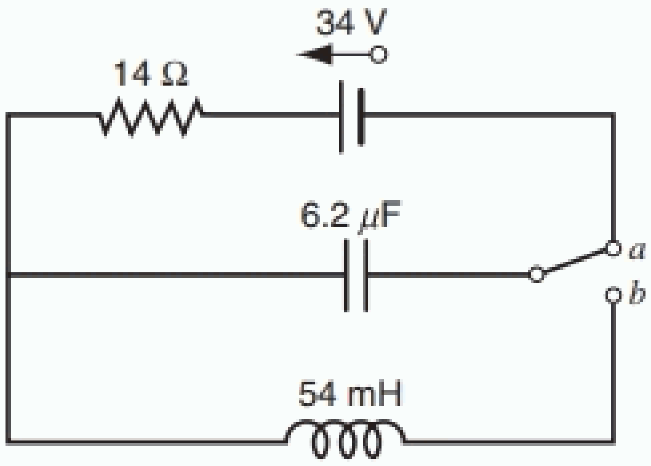


FIGURE 36-22. Exercise 44.

45. An *LC* circuit has an inductance of 3.0 mH and a capacitance of 10 μF. Calculate (*a*) the angular frequency and (*b*) the period of oscillation. (*c*) At time $t = 0$ the capacitor is charged to 200 μC, and the current is zero. Sketch roughly the charge on the capacitor as a function of time.

46. An inductor is connected across a capacitor whose capacitance can be varied by turning a knob. We wish to make the frequency of the *LC* oscillations vary linearly with the angle of rotation of the knob, going from 200 to 400 kHz as the knob turns through 180°. If $L = 1.0$ mH, plot *C* as a function of angle for the 180° rotation.

47. (*a*) In an oscillating *LC* circuit, in terms of the maximum charge on the capacitor, what value of charge is present when the energy in the electric field is one-half that in the magnetic field? (*b*) What fraction of a period must elapse following the time the capacitor is fully charged for this condition to arise?

48. In an *LC* circuit, $L = 24.8$ mH and $C = 7.73$ μF. At time $t = 0$, the current is 9.16 mA, the charge on the capacitor is 3.83 μC, and the capacitor is charging. (*a*) What is the total energy in the circuit? (*b*) What is the maximum charge on the capacitor? (*c*) What is the maximum current? (*d*) If the charge on the capacitor is given by $q = q_m \cos(\omega t + \phi)$, what is the phase angle ϕ? (*e*) Suppose the data are the same, except that the capacitor is discharging at $t = 0$. What then is the phase angle ϕ?

49. A variable capacitor with a range from 10 to 365 pF is used with a coil to form a variable-frequency *LC* circuit to tune the input to a radio. (*a*) What ratio of maximum to minimum frequencies may be tuned with such a capacitor? (*b*) If this capacitor is to tune from 0.54 to 1.60 MHz, the ratio computed in (*a*) is too large. By adding a capacitor in parallel to the variable capacitor this range may be adjusted. How large should this capacitor be and what inductance should be chosen in order to tune the desired range of frequencies?

50. In Fig. 36-23, the 900-μF capacitor is initially charged to 100 V and the 100-μF capacitor is uncharged. Describe in detail how one might charge the 100-μF capacitor to 300 V by manipulating switches S_1 and S_2.

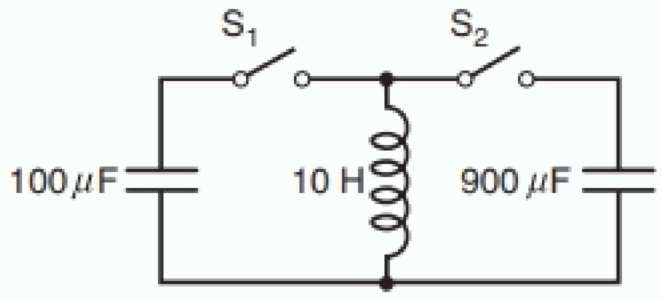


FIGURE 36-23. Exercise 50.

51. In an oscillating *LC* circuit, $L = 3.0$ mH and $C = 2.7$ μF. At $t = 0$, the charge on the capacitor is zero and the current is 2.0 A. (*a*) What is the maximum charge that will appear on the capacitor? (*b*) In terms of the period *T* of oscillation, how much time will elapse after $t = 0$ until the energy stored in the capacitor will be increasing at its greatest rate? (*c*) What is this greatest rate at which energy flows into the capacitor?

36-7 Damped and Forced Oscillations

52. A single-loop circuit consists of a 7.22-Ω resistor, a 12.3-H inductor, and a 3.18-μF capacitor. Initially, the capacitor has a charge of 6.31 μC and the current is zero. Calculate the charge on the capacitor *N* complete cycles later for $N = 5$, 10, and 100.

53. In a damped *LC* circuit, find the time required for the maximum energy present in the capacitor during one oscillation to fall to one-half of its initial value. Assume that $q = q_m$ at $t = 0$.

54. How much resistance R should be connected to an inductor $L = 220$ mH and capacitor $C = 12$ μF in series in order that the maximum charge on the capacitor decay to 99% of its initial value in 50 cycles?

55. A circuit has $L = 12.6$ mH and $C = 1.15$ μF. How much resistance must be inserted in the circuit to reduce the (undamped) resonant frequency by 0.01%?

PROBLEMS

1. A long, thin solenoid can be bent into a ring to form a toroid. Show that if the solenoid is long and thin enough, the equation for the inductance of a toroid (Eq. 36-10) is equivalent to that for a solenoid of the appropriate length (Eq. 36-7).

2. A wide copper strip of width W is bent into a piece of slender tubing of radius R with two plane extensions, as shown in Fig. 36-24. A current i flows through the strip, distributed uniformly over its width. In this way a "one-turn solenoid" has been formed. (*a*) Derive an expression for the magnitude of the magnetic field $\vec{\mathbf{B}}$ in the tubular part (far away from the edges). (Hint: Assume that the field outside this one-turn solenoid is negligibly small.) (*b*) Find also the inductance of this one-turn solenoid, neglecting the two plane extensions.

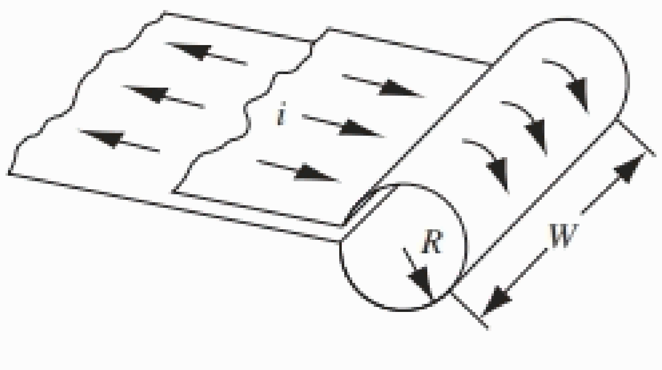


FIGURE 36-24. Problem 2.

3. Two long, parallel wires, each of radius a, whose centers are a distance d apart carry equal currents in opposite directions. Show that, neglecting the flux within the wires themselves, the inductance of a length l of such a pair of wires is given by

$$L = \frac{\mu_0 l}{\pi} \ln \frac{d - a}{a}.$$

See Sample Problem 33-4. (Hint: Calculate the flux through a rectangle of which the wires form two opposite sides.)

4. Two long, parallel copper wires (diameter = 2.60 mm) carry currents of 11.3 A in opposite directions. (*a*) If their centers are 21.8 mm apart, calculate the flux per meter of wire that exists in the space between the axes of the wires. (*b*) What fraction of this flux lies inside the wires, and therefore, what is the fractional error made in ignoring this flux in calculating the inductance of two parallel wires? See Problem 3. (*c*) Repeat the calculations of (*a*) for parallel currents.

5. Find the inductance of the coaxial cable of Fig. 36-7 directly from Eq. 36-4. (Hint: Calculate the flux through a rectangular surface, perpendicular to $\vec{\mathbf{B}}$, of length l and width $b - a$.)

6. In Fig. 36-25 the component in the upper branch is an ideal 3.0-A fuse. It has zero resistance as long as the current through it remains less than 3.0 A. If the current reaches 3.0 A, it "blows" and thereafter it has infinite resistance. Switch S is closed at time $t = 0$. (*a*) When does the fuse blow? (*b*) Sketch a graph of the current i through the inductor as a function of time. Mark the time at which the fuse blows.

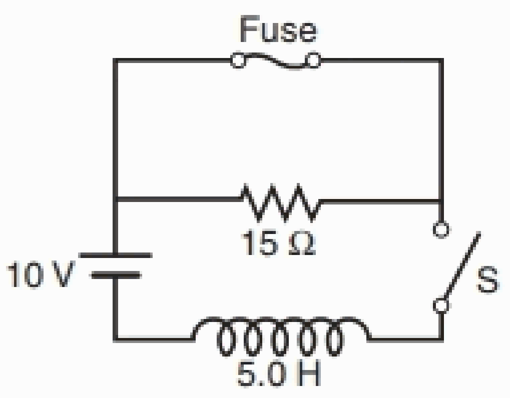


FIGURE 36-25. Problem 6.

7. Show that the inductive time constant τ_L can also be defined as the time that would be required for the current in an *LR* circuit to reach its equilibrium value if it continued to increase at its initial rate.

8. The coil of a superconducting electromagnet used for nuclear magnetic resonance investigations has an inductance of 152 H and carries a current of 32 A. The coil is immersed in liquid helium, which has a latent heat of vaporization of 85 J/mol. (*a*) Calculate the energy in the magnetic field of the coil. (*b*) Find the mass of helium that is boiled off if the superconductor is quenched and thereby suddenly develops a finite resistance.

9. (*a*) Find an expression for the energy density as a function of the radial distance r for a toroid of rectangular cross section. (*b*) Integrating the energy density over the volume of the toroid, calculate the total energy stored in the field of the toroid. (*c*) Using Eq. 36-10, evaluate the energy stored in the toroid directly from the inductance and compare with (*b*).

10. Prove that, after switch S in Fig. 36-4 is thrown from a to b, all the energy stored in the inductor ultimately appears as internal energy in the resistor.

11. A long wire carries a current i uniformly distributed over a cross section of the wire. (*a*) Show that the magnetic energy of a length l stored within the wire equals $\mu_0 i^2 l/16\pi$. (Why does it not depend on the wire diameter?) (*b*) Show that the inductance for a length l of the wire associated with the flux inside the wire is $\mu_0 l/8\pi$.

12. The resonant frequency of a series circuit containing inductance L_1 and capacitance C_1 is ω_0. A second series circuit, containing inductance L_2 and capacitance C_2, has the same resonant frequency. In terms of ω_0, what is the resonant frequency of a series circuit containing all four of these elements? Neglect resistance. (Hint: Use the formulas for equivalent capacitance and equivalent inductance.)

13. Three identical inductors L and two identical capacitors C are connected in a two-loop circuit as shown in Fig. 36-26. (*a*) Suppose the currents are as shown in Fig. 36-26*a*. What is the current in the middle inductor? Write down the loop equations and show that they are satisfied provided that the current oscillates with angular frequency $\omega = 1/\sqrt{LC}$. (*b*) Now suppose the currents are as shown in Fig. 36-26*b*. What is the current in the middle inductor? Write down the loop equations and show that they are satisfied provided the current oscillates with angular frequency $\omega = 1/\sqrt{3LC}$. (*c*) In view of the fact that the circuit can oscillate at two different frequencies, show that it is not possible to replace this two-loop circuit by an equivalent single-loop LC circuit.

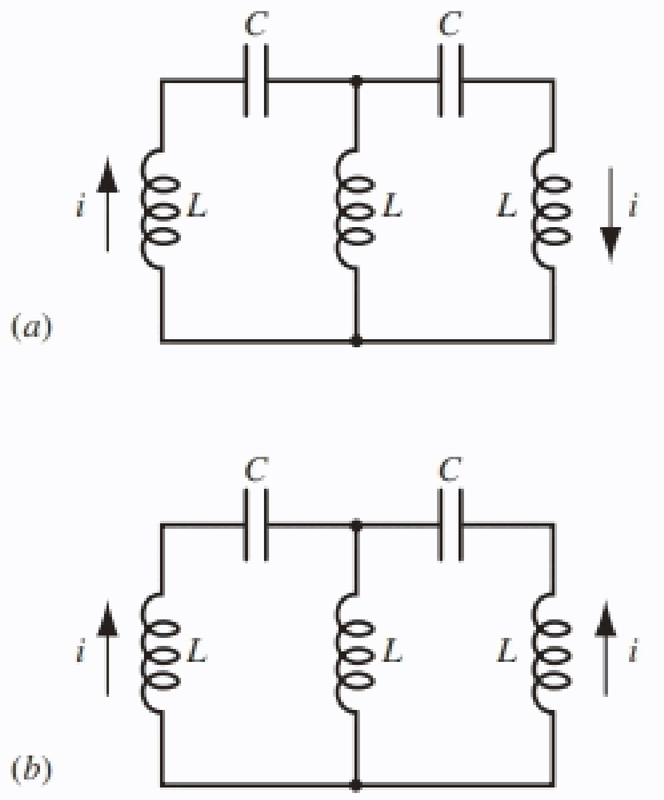


FIGURE 36-26. Problem 13.

14. (*a*) By direct substitution of Eq. 36-40 into Eq. 36-39, show that $\omega' = \sqrt{\omega^2 - (R/2L)^2}$. (*b*) By what fraction does the frequency of oscillation shift when the resistance is increased from 0 to 100 Ω in a circuit with $L = 4.4$ H and $C = 7.3$ μF?

15. In a damped LC circuit, show that the fraction of the energy lost per cycle of oscillation, $\Delta U/U$, is given to a close approximation by $2\pi R/\omega L$. The quantity $\omega L/R$ is often called the Q of the circuit (for "quality"). A "high-Q" circuit has low resistance and a low fractional energy loss per cycle $(=2\pi/Q)$.

16. Suppose that in a damped LC circuit the amplitude of the charge oscillations drops to one-half its initial value after n cycles. Show that the fractional reduction in the frequency of resonance, caused by the presence of the resistor, is given to a close approximation by

$$\frac{\omega - \omega'}{\omega} = \frac{0.0061}{n^2},$$

which is independent of L, C, or R.

COMPUTER PROBLEMS

1. A real cylindrical solenoid has length l, diameter $d = 0.10l$, and n turns per unit length. Assuming that the magnetic field at off-axis points inside the solenoid is accurately described by the exact expression along the axis (see Eq. 33-27), numerically calculate the inductance of the solenoid. Compare your answer to the ideal solenoid approximation. In this numerical result we have neglected variation of the field at points off the axis. Does that result in an answer which is too large or too small?

2. Calculate the inductance of a flat wire loop of radius R. Assume the wire has a radius $r = 0.010R$, and that the contribution to the inductance from the magnetic field inside the wire is negligible.

CHAPTER 37

ALTERNATING CURRENT CIRCUITS

Circuits involving alternating currents *(commonly abbreviated AC) are used in electric power distribution systems, in radio, television, and other communication devices, and in a wide variety of electric motors. The designation "alternating" means that the current changes direction, alternating periodically from one direction to the other. Generally we work with currents that vary sinusoidally with time; however, as we have seen previously in the case of wave motion, complex waveforms can be viewed as combinations of sinusoidal waves (through Fourier analysis), and by analogy we can understand the behavior of circuits having currents of arbitrary time dependence by first understanding the behavior of circuits having currents that vary sinusoidally with time.*

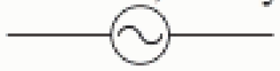
In this chapter we study the behavior of simple circuits containing resistors, inductors, and capacitors when a sinusoidally varying source of emf is present.

37-1 ALTERNATING CURRENTS

In Section 36-7 we considered a series *RLC* circuit by focusing on its behavior at resonance, where the frequency of the driving emf is equal to the natural frequency of the *LC* oscillator. Here we consider the same circuit in which the driving frequency can be far from resonance. Typically, *LC* circuits have resonance frequencies in the kHz to MHz range, whereas the driving emf is usually the 60 Hz supplied by electrical power companies in the United States. Our general approach in this chapter is valid for any driving frequency and includes resonance as a special case.

We can consider the driving emf to originate from a generator of the type discussed in Chapter 34 (see Fig. 34-13). The generator produces an emf that varies sinusoidally (Fig. 34-14), which we write as

$$\mathcal{E}(t) = \mathcal{E}_{\mathrm{m}} \sin \omega t, \tag{37-1}$$

where $\mathcal{E}_{\mathrm{m}}$ is the amplitude of the varying emf and ω is its angular frequency (in rad/s), related to the frequency f (in Hz) by $\omega = 2\pi f$. In a circuit, the symbol for this source of time-varying emf is . As the emf varies between positive and negative values in each cycle, the current in the circuit changes direction. We call such a circuit an *alternating current* (AC) circuit.

Our goal in this chapter is to understand the result of applying an alternating emf, of the form of Eq. 37-1, to a circuit containing resistive, inductive, and capacitive elements. There are many ways these elements can be connected in a circuit; as an example of the analysis of AC circuits, we consider in this chapter the series *RLC* circuit shown in Fig. 37-1, in which a resistor *R*, inductor *L*, and capacitor *C* are connected in series across an alternating emf of the form of Eq. 37-1.

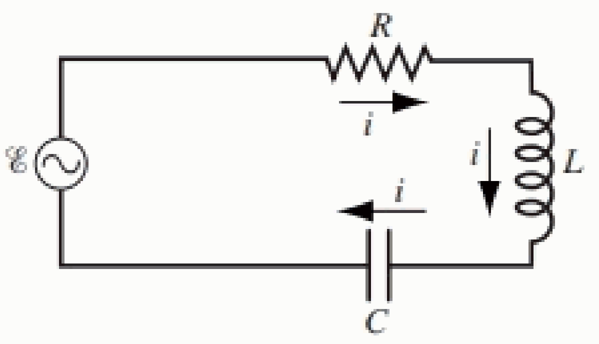


FIGURE 37-1. A single-loop circuit, consisting of a resistor, an inductor, and a capacitor. A generator supplies a source of alternating emf that establishes an alternating current.

For a short time after the emf is initially applied to the circuit, the current varies erratically with time. These variations, called *transients,* rapidly die away, after which we find that *the current varies sinusoidally with the same angular frequency as the source of emf.* We assume that we examine the circuit after it has settled into this condition, in which the current can be written as

$$i = i_{\mathrm{m}} \sin(\omega t - \phi), \tag{37-2}$$

where i_{m} is the *current amplitude* (the maximum magnitude of the current) and ϕ is a phase constant or phase angle that indicates the phase relationship between $\mathscr{E}$ and i. (Note that we have assumed a phase constant of 0 in Eq. 37-1 for the emf. Note also that we write the phase constant in Eq. 37-2 with a negative sign; this choice is customary in discussing the phase relationship between the current and the emf.) The angular frequency ω in Eq. 37-2 is the same as that in Eq. 37-1.

We assume that $\mathscr{E}_{\mathrm{m}}$, ω, R, L, and C are known. The goal of our calculation is to find i_{m} and ϕ, so that Eq. 37-2 completely characterizes the current. We use a general method for the series *RLC* circuit; a similar procedure can be used to analyze more complicated circuits (containing elements in various series and parallel combinations). It can also be applied to nonsinusoidal emfs, because more complicated emfs can be written in terms of sinusoidal emfs using the techniques of Fourier analysis (see Section 18-7), and the resultant current can similarly be considered to be the superposition of many terms of the form of Eq. 37-2. Understanding the series *RLC* circuit driven by a sinusoidal emf is therefore essential to understanding time-dependent behavior in all circuits.

37-2 THREE SEPARATE ELEMENTS

Before analyzing the circuit of Fig. 37-1, it is helpful to discuss the response of each of the three elements separately to an alternating current of the form of Eq. 37-2. We assume that we deal with ideal elements; for instance, the inductor has only inductance and no resistance or capacitance.

A Resistive Element

Figure 37-2*a* shows a resistor in a section of a circuit in which a current i (given by Eq. 37-2) has been established by means not shown in the figure. Defining ΔV_R $(= V_a - V_b)$ as the potential difference across the resistor, we can write

$$\Delta V_R = iR = i_{\mathrm{m}} R \sin(\omega t - \phi). \tag{37-3}$$

Comparison of Eqs. 37-2 and 37-3 shows that the time-varying quantities ΔV_R and i are *in phase:* they reach their maximum values at the same time. This phase relationship is illustrated in Fig. 37-2*b*.

Figure 37-2*c* shows another way of looking at the situation. It is called a *phasor diagram,* in which the phasors, represented by the open arrows, rotate counterclockwise with an angular frequency ω about the origin. The phasors have the following properties. (1) The length of a phasor is proportional to the *maximum* value of the alternating quantity involved: for the potential difference, $(\Delta V_R)_{\max} = i_{\mathrm{m}} R$ from Eq. 37-3, and for the current, i_{m} from Eq. 37-2. (2) The projection of a phasor on the vertical axis gives the *instantaneous* value of the alternating quantity involved. The arrows on the vertical axis represent the time-varying quantities ΔV_R and i, as in Eqs. 37-2 and 37-3, respectively. That ΔV_R and i are in phase follows from the fact that their phasors lie along the same line in Fig. 37-2*c*.

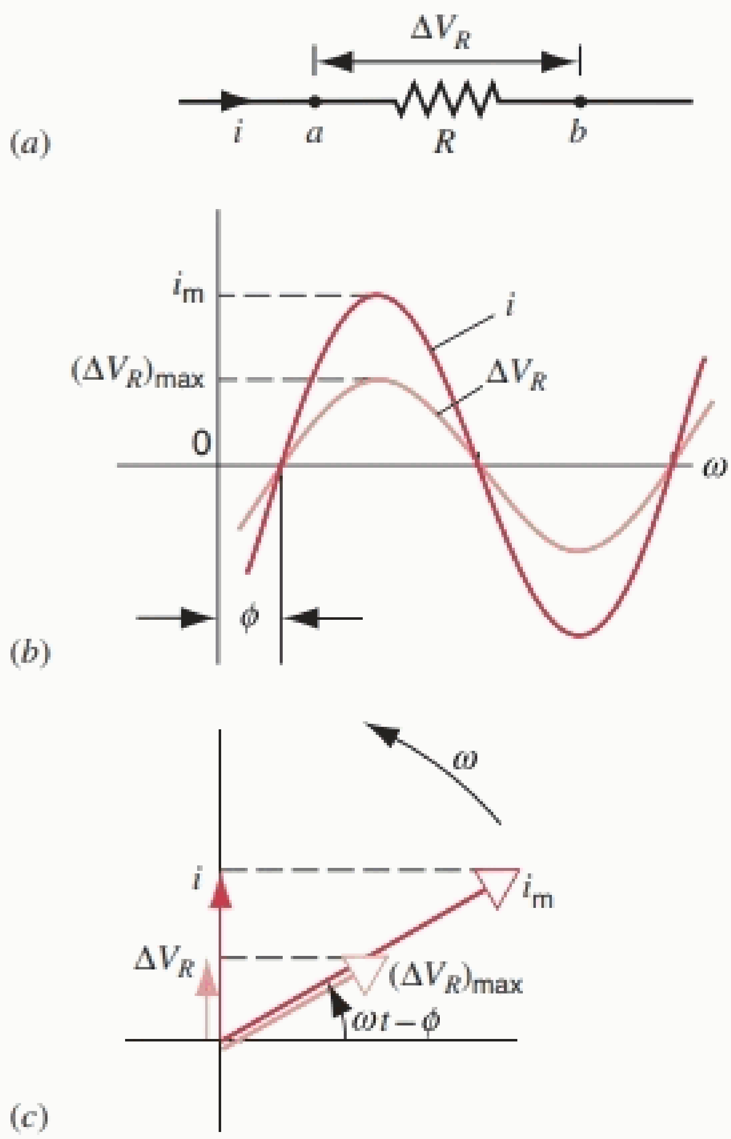


FIGURE 37-2. (*a*) A resistor in an AC circuit. (*b*) The current and the potential difference across the resistor are in phase. (*c*) A phasor diagram representing the current and potential difference.

The phasor diagram is very similar to Fig. 17-14, in which we made the connection between uniform circular motion and simple harmonic motion. You may recall that the projection on any axis of the position of a particle moving in uniform circular motion gives a displacement that varies sinusoidally, in analogy with simple harmonic motion. Here as the phasors rotate, their projections on the vertical axis give a sinusoidally varying current or voltage. Follow the rotation of the phasors in Fig. 37-2*c* and convince yourself that this phasor diagram completely and correctly describes Eqs. 37-2 and 37-3.

An Inductive Element

Figure 37-3*a* shows a portion of a circuit containing only an inductive element. The potential difference ΔV_L $(= V_a - V_b)$ across the inductor is related to the current by Eq. 36-2:

$$\Delta V_L = L\frac{di}{dt} = L i_{\mathrm{m}} \omega \cos(\omega t - \phi), \tag{37-4}$$

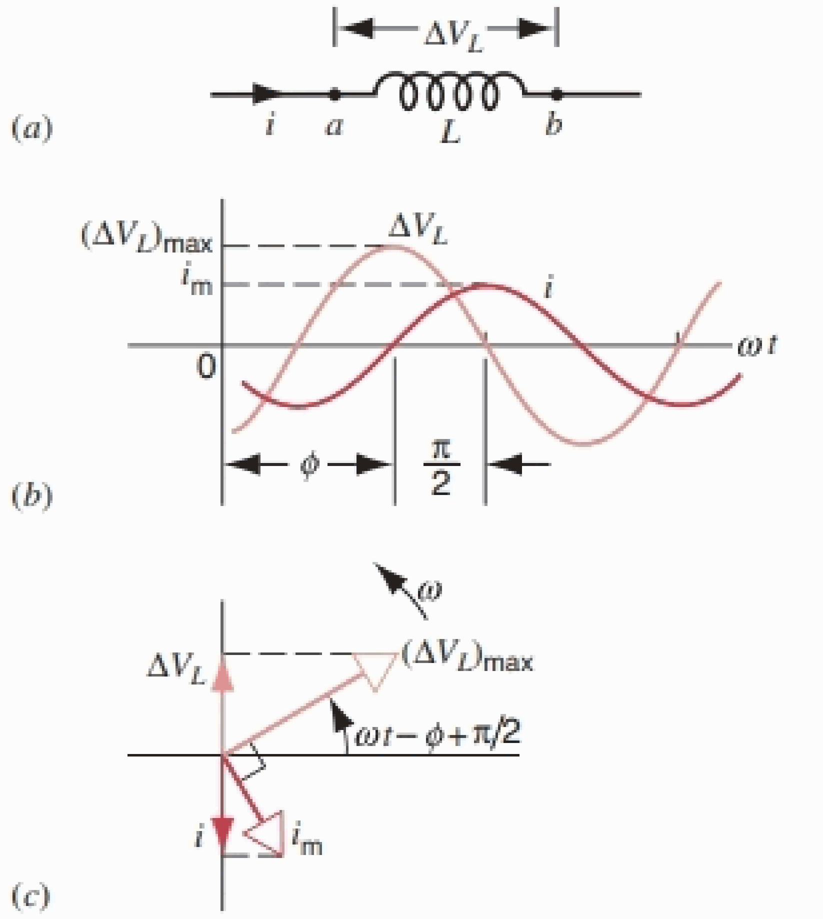


FIGURE 37-3. (*a*) An inductor in an AC circuit. (*b*) The current lags the potential difference across the inductor by 90°. (*c*) A phasor diagram representing the current and potential difference.

which we obtain by taking the derivative of Eq. 37-2 for the current. The trigonometric identity $\cos\theta = \sin(\theta + \pi/2)$ allows us to write Eq. 37-4 as

$$\Delta V_L = L i_{\mathrm{m}} \omega \sin(\omega t - \phi + \pi/2). \qquad (37\text{-}5)$$

Comparison of Eqs. 37-2 and 37-5 shows that the time-varying quantities ΔV_L and i are not in phase; they are one-quarter cycle out of phase, with *i behind ΔV_L*. It is commonly said that the current *lags* the potential difference by 90° in an inductor. We show this in Fig. 37-3*b*, which is a plot of Eqs. 37-2 and 37-5. Note that, as time goes on, i reaches its maximum *after* ΔV_L does, by one-quarter cycle.

This phase relationship between i and ΔV_L is indicated in the phasor diagram of Fig. 37-3*c*. As the phasors rotate counterclockwise, it is clear that the i phasor follows (that is, *lags*) the ΔV_L phasor by one-quarter cycle.

In analyzing AC circuits, it is convenient to define the *inductive reactance X_L*:

$$X_L = \omega L, \qquad (37\text{-}6)$$

in terms of which we can rewrite Eq. 37-5 as

$$\Delta V_L = i_{\mathrm{m}} X_L \sin(\omega t - \phi + \pi/2). \qquad (37\text{-}7)$$

Comparing Eqs. 37-3 and 37-7, we see that the SI unit for X_L must be the same as that of R—namely, the ohm. This can be seen directly by comparing Eq. 37-6 with the expression for the inductive time constant, $\tau_L = L/R$. Even though both are measured in ohms, a reactance is not the same as a resistance.

The maximum value of ΔV_L is, from Eq. 37-7,

$$(\Delta V_L)_{\max} = i_{\mathrm{m}} X_L. \qquad (37\text{-}8)$$

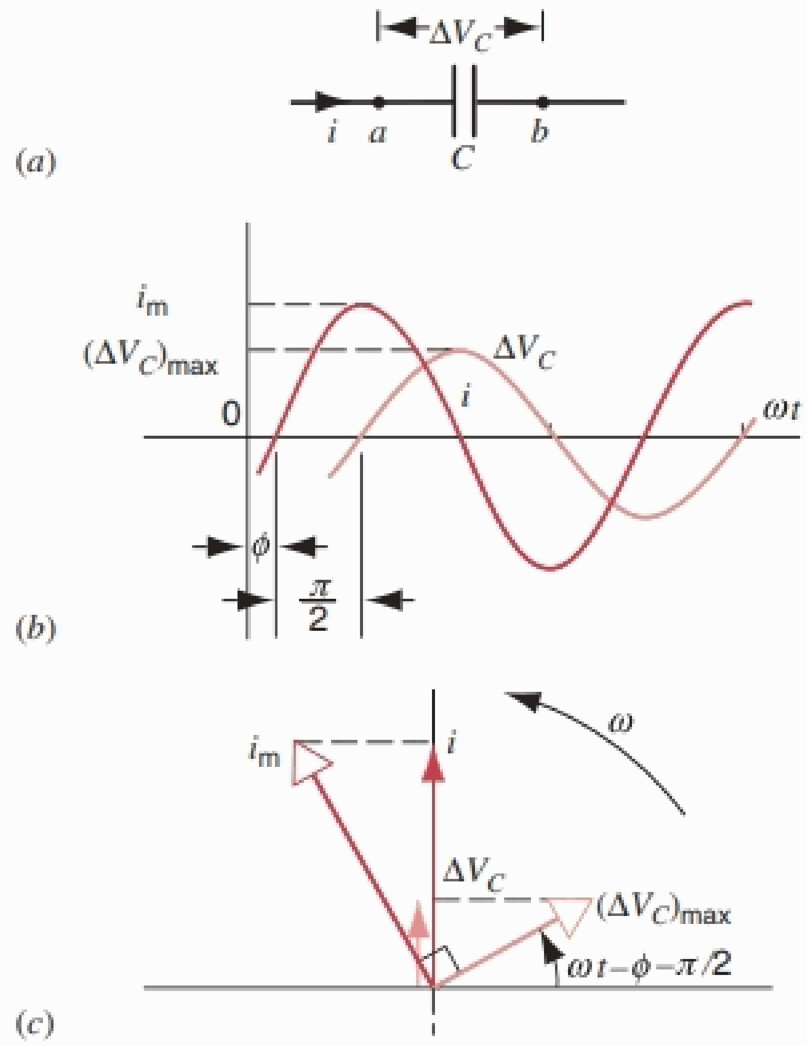


FIGURE 37-4. (*a*) A capacitor in an AC circuit. (*b*) The current leads the potential difference across the capacitor by 90° (*c*) A phasor diagram representing the current and potential difference.

A Capacitive Element

Figure 37-4*a* shows a portion of a circuit containing only a capacitive element. Again, a current i given by Eq. 37-2 has been established by means not shown.* Let the charge on the left-hand plate be q, so that a positive current into that plate gives an increase in q; that is, $i = dq/dt$ implies that $dq > 0$ when $i > 0$. The potential difference $\Delta V_C\ (= V_a - V_b)$ across the capacitor is given by

$$\Delta V_C = \frac{q}{C} = \frac{\int i\,dt}{C}. \qquad (37\text{-}9)$$

Integrating the current i given by Eq. 37-2, we find

$$\Delta V_C = -\frac{i_{\mathrm{m}}}{\omega C}\cos(\omega t - \phi)$$

$$= \frac{i_{\mathrm{m}}}{\omega C}\sin(\omega t - \phi - \pi/2), \qquad (37\text{-}10)$$

where we have used the trigonometric identity $\cos\theta = -\sin(\theta - \pi/2)$.

---

* It may at first be difficult to think of a capacitor as a part of a current-carrying circuit; clearly charge does not flow through the capacitor. It may be helpful to consider the flow of charge in this way: the current i brings charge q to the left-hand plate of the capacitor, so a charge $-q$ must flow to the right-hand plate from whatever circuit is beyond the capacitor to the right. This flow of charge $-q$ from right to left is entirely equivalent to a flow of charge $+q$ from left to right, which is identical to the current on the left side of the capacitor. Thus a current on one side of the capacitor can appear on the other side, even though there is no conducting path between the two plates!

TABLE 37-1 Phase and Amplitude Relations for Alternating Currents and Voltages

Circuit Element	*Symbol*	*Impedance*[a]	*Phase of the Current*	*Amplitude Relation*
Resistor	R	R	In phase with ΔV_R	$(\Delta V_R)_{\max} = i_m R$
Inductor	L	X_L	Lags ΔV_L by 90°	$(\Delta V_L)_{\max} = i_m X_L$
Capacitor	C	X_C	Leads ΔV_C by 90°	$(\Delta V_C)_{\max} = i_m X_C$

Many students have remembered the phase relations from: "ELI the ICE man." Here L and C stand for inductance and capacitance; E stands for voltage and I for current. Thus in an inductive circuit (ELI) the current (I) lags the voltage (E), and in a capacitive circuit (ICE) the current (I) leads the voltage (E).

[a] *Impedance* is a general term that includes both resistance and reactance.

Comparing Eqs. 37-2 and 37-10, we see that i and ΔV_C are 90° out of phase, with i *ahead of* ΔV_C. Figure 37-4*b* shows i and ΔV_C plotted as functions of the time; note that i reaches its maximum one-quarter cycle or 90° *before* ΔV_C. Equivalently, we may say that the current *leads* the potential difference by 90° in a capacitor.

The phase relationship is shown in the phasor diagram of Fig. 37-4*c*. As the phasors rotate counterclockwise, it is clear that the i phasor leads the ΔV_C phasor by one-quarter cycle.

In analogy with the inductive reactance, it is convenient to define the *capacitive reactance* X_C:

$$X_C = \frac{1}{\omega C}, \qquad (37\text{-}11)$$

in terms of which we can rewrite Eq. 37-10 as

$$\Delta V_C = i_m X_C \sin(\omega t - \phi - \pi/2). \qquad (37\text{-}12)$$

Comparing Eqs. 37-3 and 37-12, we see that the unit of X_C must also be the ohm. This conclusion also follows by comparing Eq. 37-11 with the expression $\tau_C = RC$ for the capacitive time constant.

The maximum value of ΔV_C is, from Eq. 37-12,

$$(\Delta V_C)_{\max} = i_m X_C. \qquad (37\text{-}13)$$

Table 37-1 summarizes the results derived for the three individual circuit elements.

SAMPLE PROBLEM 37-1. In Fig. 37-3*a* let $L = 230$ mH, $f = 60$ Hz, and $(\Delta V_L)_{\max} = 36$ V. (*a*) Find the inductive reactance X_L. (*b*) Find the current amplitude in the circuit.

Solution (*a*) From Eq. 37-6,

$$X_L = \omega L = 2\pi f L = (2\pi)(60\text{ Hz})(230 \times 10^{-3}\text{ H}) = 87\ \Omega.$$

(*b*) From Eq. 37-8, the current amplitude is

$$i_m = \frac{(\Delta V_L)_{\max}}{X_L} = \frac{36\text{ V}}{87\ \Omega} = 0.41\text{ A}.$$

We see that, although a reactance is not a resistance, the inductive reactance plays the same role for an inductor that the resistance does for a resistor. Note that, if you doubled the frequency, the inductive reactance would double and the current amplitude would be cut in half. We can also understand this physically: to get the same value of ΔV_L, you must change the current at the same rate ($\Delta V_L = L\,di/dt$). If the frequency doubles, you cut the time of change in half so the maximum current is also cut in half. To sum up: for inductors, the higher the frequency, the higher the reactance.

SAMPLE PROBLEM 37-2. In Fig. 37-4*a*, let $C = 15\ \mu$F, $f = 60$ Hz, and $(\Delta V_C)_{\max} = 36$ V. (*a*) Find the capacitive reactance X_C. (*b*) Find the current amplitude in this circuit.

Solution (*a*) From Eq. 37-11, we have

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi f C} = \frac{1}{(2\pi)(60\text{ Hz})(15 \times 10^{-6}\text{ F})} = 177\ \Omega.$$

(*b*) From Eq. 37-13, we have for the current amplitude

$$i_m = \frac{(\Delta V_C)_{\max}}{X_C} = \frac{36\text{ V}}{177\ \Omega} = 0.20\text{ A}.$$

Note that, if you doubled the frequency, the capacitive reactance would drop to half its value and the current amplitude would double. We can understand this physically: to get the same value of ΔV_C you must deliver the same charge to the capacitor ($\Delta V_C = q/C$). If the frequency doubles, then you have only half the time to deliver this charge so the maximum current must double. To sum up: for capacitors, the higher the frequency, the lower the reactance.

37-3 THE SINGLE-LOOP *RLC* CIRCUIT

Having finished our analysis of separate R, L, and C elements, we now return to the analysis of the circuit of Fig. 37-1, in which all three elements are present. The emf is given by Eq. 37-1,

$$\mathcal{E} = \mathcal{E}_m \sin \omega t,$$

and the current in the circuit has the form of Eq. 37-2,

$$i = i_m \sin(\omega t - \phi).$$

Our goal is to determine i_m and ϕ.

We start by applying the loop rule (Section 31-3) to the circuit of Fig. 37-1, obtaining $\mathcal{E} - \Delta V_R - \Delta V_L - \Delta V_C = 0$, or

$$\mathcal{E} = \Delta V_R + \Delta V_L + \Delta V_C. \qquad (37\text{-}14)$$

Equation 37-14 can be solved for the current amplitude i_m and phase ϕ using a variety of techniques: a trigonometric analysis, a graphical analysis using phasors, and a differential analysis.

Trigonometric Analysis

We have already obtained relationships between the potential difference across each element and the current through the element. Let us therefore substitute Eqs. 37-3, 37-7, and 37-12 into Eq. 37-14, from which we obtain

$$\begin{aligned}\mathcal{E}_{\mathrm{m}} \sin \omega t = {} & i_{\mathrm{m}} R \sin(\omega t - \phi) \\ & + i_{\mathrm{m}} X_L \sin(\omega t - \phi + \pi/2) \\ & + i_{\mathrm{m}} X_C \sin(\omega t - \phi - \pi/2), \quad (37\text{-}15)\end{aligned}$$

in which we have substituted Eq. 37-1 for the emf. Using trigonometric identities, Eq. 37-15 can be written

$$\begin{aligned}\mathcal{E}_{\mathrm{m}} \sin \omega t &= i_{\mathrm{m}} R \sin(\omega t - \phi) + i_{\mathrm{m}} X_L \cos(\omega t - \phi) \\ &\quad - i_{\mathrm{m}} X_C \cos(\omega t - \phi) \\ &= i_{\mathrm{m}}[R \sin(\omega t - \phi) \\ &\quad + (X_L - X_C) \cos(\omega t - \phi)]. \qquad (37\text{-}16)\end{aligned}$$

which reduces to (see Exercise 12)

$$\mathcal{E}_{\mathrm{m}} \sin \omega t = i_{\mathrm{m}} \sqrt{R^2 + (X_L - X_C)^2} \sin \omega t \quad (37\text{-}17)$$

provided we choose

$$\tan \phi = \frac{X_L - X_C}{R} = \frac{\omega L - 1/\omega C}{R}. \qquad (37\text{-}18)$$

The current amplitude is found directly from Eq. 37-17:

$$i_{\mathrm{m}} = \frac{\mathcal{E}_{\mathrm{m}}}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{\mathcal{E}_{\mathrm{m}}}{\sqrt{R^2 + (\omega L - 1/\omega C)^2}}. \qquad (37\text{-}19)$$

This completes the analysis of the series *RLC* circuit, because we have accomplished our goal of expressing the current amplitude i_{m} and phase ϕ in terms of the parameters of the circuit ($\mathcal{E}_{\mathrm{m}}$, ω, R, L, and C). Note that the phase ϕ does not depend on the amplitude $\mathcal{E}_{\mathrm{m}}$ of the applied emf. Changing $\mathcal{E}_{\mathrm{m}}$ changes i_{m} but not ϕ; the *scale* of the result changes but not its *nature.*

The quantity in the denominator of Eq. 37-19 is called the *impedance* Z of the series *RLC* circuit:

$$Z = \sqrt{R^2 + (X_L - X_C)^2}, \qquad (37\text{-}20)$$

and so Eq. 37-19 can be written

$$i_{\mathrm{m}} = \frac{\mathcal{E}_{\mathrm{m}}}{Z}, \qquad (37\text{-}21)$$

which reminds us of the relation $i = \mathcal{E}/R$ for single-loop resistive networks with steady emfs. The SI unit of impedance is evidently the ohm.

Equation 37-19 gives the current amplitude in Eq. 36-43, and Fig. 36-13 is a plot of Eq. 37-19. The current i_{m} has its maximum value when the impedance Z has its minimum value R, which occurs when $X_L = X_C$, or

$$\omega L = 1/\omega C,$$

so that

$$\omega = 1/\sqrt{LC}, \qquad (37\text{-}22)$$

which is the *resonance* condition given in Eq. 36-44. Although Eq. 37-19 is a general result valid for any driving frequency, it includes the resonance condition as a special case.

Graphical Analysis

It is instructive to use a phasor diagram to analyze the series *RLC* circuit. Figure 37-5*a* shows a phasor representing the current. It has length i_{m}, and its projection on the vertical axis is $i_{\mathrm{m}} \sin(\omega t - \phi)$, which is the time-varying current i. In Fig. 37-5*b* we have drawn phasors representing the individual potential differences across R, L, and C. Note their maximum values and time-varying projections on the vertical axis. Be sure to note that the phases are in agreement with our conclusions from Section 37-2: ΔV_R is in phase with the current, ΔV_L *leads* the current by 90°, and ΔV_C *lags* the current by 90°.

In accordance with Eq. 37-14, the *algebraic* sum of the (instantaneous) projections of ΔV_R, ΔV_L, and ΔV_C on the

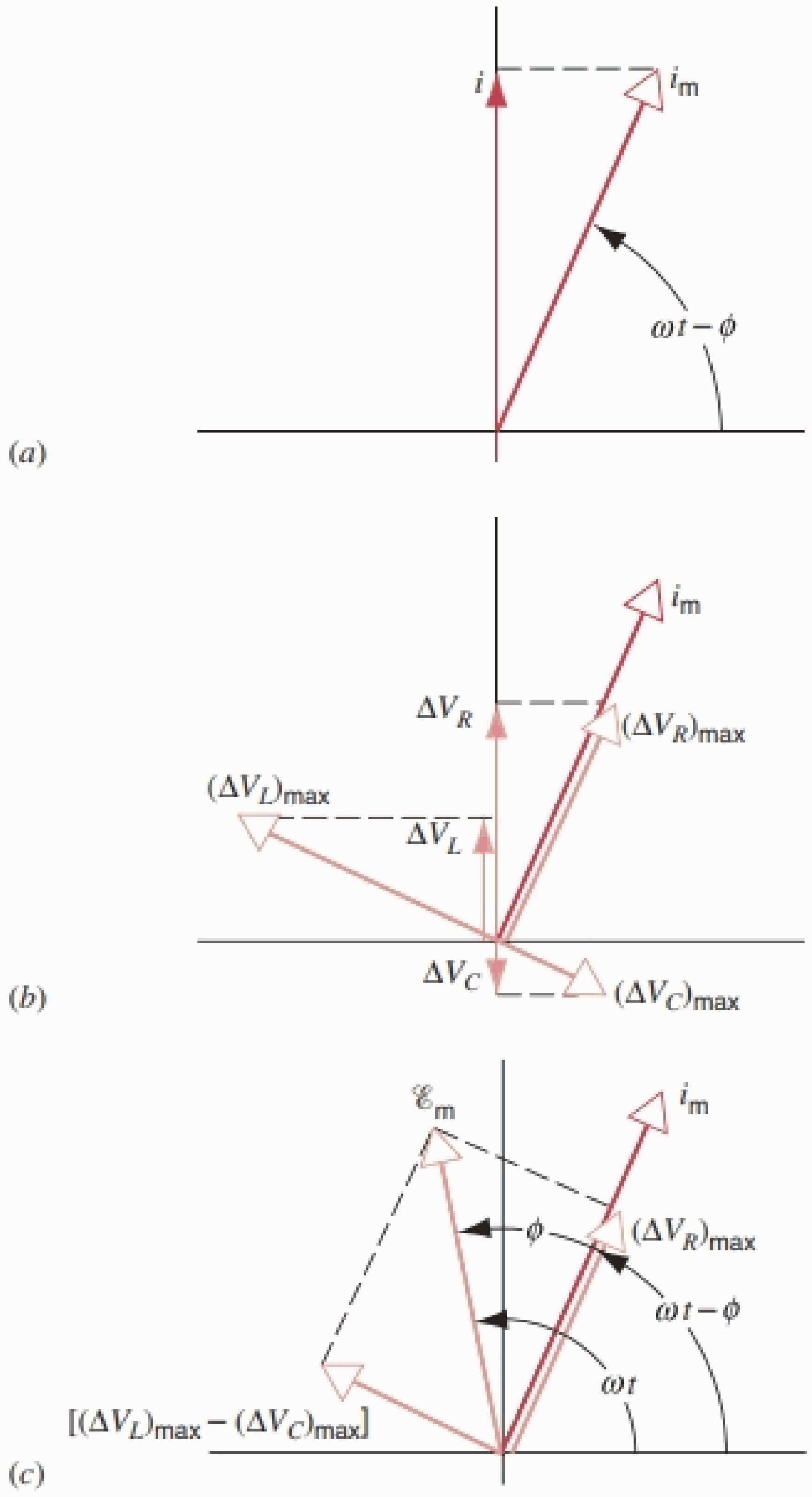


FIGURE 37-5. (*a*) A phasor representing the alternating current in the *RLC* circuit of Fig. 37-1. (*b*) Phasors representing the potential differences across the resistor, capacitor, and inductor. Note their phase differences with respect to the current. (*c*) A phasor representing the alternating emf has been added.

vertical axis gives the (instantaneous) value of $\mathcal{E}$. On the other hand, we assert that the *vector* sum of the phasor amplitudes $(\Delta V_R)_{\max}$, $(\Delta V_L)_{\max}$, and $(\Delta V_C)_{\max}$ yields a phasor whose amplitude is the $\mathcal{E}_m$ of Eq. 37-1. The projection of $\mathcal{E}_m$ on the vertical axis is the time-varying $\mathcal{E}$ of Eq. 37-1; that is, it is $\Delta V_R + \Delta V_L + \Delta V_C$ as Eq. 37-14 asserts. In vector operations, the (algebraic) sum of the projections of any number of vectors on a given straight line is equal to the projection on that line of the (vector) sum of those vectors.

In Fig. 37-5*c*, we have first formed the vector sum of $(\Delta V_L)_{\max}$ and $(\Delta V_C)_{\max}$, which is the phasor $(\Delta V_L)_{\max} - (\Delta V_C)_{\max}$. Next we form the vector sum of this phasor with $(\Delta V_R)_{\max}$. Because these two phasors are at right angles, the amplitude of their sum, which is the amplitude of the phasor $\mathcal{E}_m$, is

$$\begin{aligned}\mathcal{E}_m &= \sqrt{[(\Delta V_R)_{\max}]^2 + [(\Delta V_L)_{\max} - (\Delta V_C)_{\max}]^2}\\ &= \sqrt{(i_m R)^2 + (i_m X_L - i_m X_C)^2}\\ &= i_m\sqrt{R^2 + (X_L - X_C)^2},\end{aligned} \tag{37-23}$$

using Eqs. 37-3, 37-8, and 37-13 to replace the phasor amplitudes. Equation 37-23 is identical with Eq. 37-19, which we obtained from the trigonometric analysis.

As shown in Fig. 37-5*c*, ϕ is the angle between the i_m and $\mathcal{E}_m$ phasors, and we see from the figure that

$$\begin{aligned}\tan\phi &= \frac{(\Delta V_L)_{\max} - (\Delta V_C)_{\max}}{(\Delta V_R)_{\max}}\\ &= \frac{i_m(X_L - X_C)}{i_m R} = \frac{X_L - X_C}{R},\end{aligned} \tag{37-24}$$

which is identical with Eq. 37-18.

We drew Fig. 37-5*b* arbitrarily with $X_L > X_C$; that is, we assumed the circuit of Fig. 37-1 to be more inductive than capacitive. For this assumption, i_m *lags* $\mathcal{E}_m$ (although not by as much as one-quarter cycle, as it did in the purely inductive element shown in Fig. 37-3). The phase angle ϕ in Eq. 37-23 (and thus in Eq. 37-2) is positive but less than $+90°$.

If, on the other hand, we had $X_C > X_L$, the circuit would be more capacitive than inductive and i_m would lead $\mathcal{E}_m$ (although not by as much as one-quarter cycle, as it did in the purely capacitive element shown in Fig. 37-4). Consistent with this change from lagging to leading, the angle ϕ in Eq. 37-23 (and thus in Eq. 37-2) would automatically become negative.

Another way of interpreting the resonance condition makes use of the phasor diagram of Fig. 37-5. At resonance, $X_L = X_C$ and, according to Eq. 37-24, $\phi = 0$. In this case, the phasors $(\Delta V_L)_{\max}$ and $(\Delta V_C)_{\max}$ in Fig. 37-5 are equal and opposite, and so i_m is in phase with $\mathcal{E}_m$.

Once again, keep in mind that, although the techniques we have demonstrated here are valid for *any* AC circuit, the results hold *only* for the series *RLC* circuit. Furthermore, remember that we are examining the circuit only in the steady-state situation, after the short-lived transient variations have died away.

SAMPLE PROBLEM 37-3. In Fig. 37-1, let $R = 160\ \Omega$, $C = 15\ \mu F$, $L = 230$ mH, $f = 60$ Hz, and $\mathcal{E}_m = 36$ V. Find (*a*) the inductive reactance X_L, (*b*) the capacitive reactance X_C, (*c*) the impedance Z for the circuit, (*d*) the current amplitude i_m, and (*e*) the phase constant ϕ.

Solution (*a*) $X_L = 87\ \Omega$, as in Sample Problem 37-1. (*b*) $X_C = 177\ \Omega$, as in Sample Problem 37-2. Note that $X_C > X_L$, so the circuit is more capacitive than inductive. (*c*) From Eq. 37-20,

$$\begin{aligned}Z &= \sqrt{R^2 + (X_L - X_C)^2}\\ &= \sqrt{(160\ \Omega)^2 + (87\ \Omega - 177\ \Omega)^2} = 184\ \Omega.\end{aligned}$$

(*d*) From Eq. 37-21,

$$i_m = \frac{\mathcal{E}_m}{Z} = \frac{36\ \text{V}}{184\ \Omega} = 0.196\ \text{A}.$$

(*e*) From Eq. 37-24 we have

$$\tan\phi = \frac{X_L - X_C}{R} = \frac{87\ \Omega - 177\ \Omega}{160\ \Omega} = -0.563.$$

Thus we have

$$\phi = \tan^{-1}(-0.563) = -29.4°.$$

A negative phase constant is appropriate for a capacitive load, as can be inferred from Table 37-1 and Fig. 37-5.

SAMPLE PROBLEM 37-4. (*a*) What is the resonance frequency in Hz of the circuit of Sample Problem 37-3? (*b*) What is the current amplitude at resonance?

Solution (*a*) From Eq. 37-22,

$$\omega = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{(0.23\ \text{H})(15 \times 10^{-6}\ \text{F})}} = 538\ \text{rad/s}.$$

Then

$$f = \frac{\omega}{2\pi} = 86\ \text{Hz}.$$

(*b*) At resonance, $X_L = X_C$, and so $Z = R$. From Eq. 37-21,

$$i_m = \frac{\mathcal{E}_m}{R} = \frac{36\ \text{V}}{160\ \Omega} = 0.23\ \text{A}.$$

The 60-Hz frequency of Sample Problem 37-3 is fairly close to resonance.

Differential Analysis (Optional)

With $\Delta V_C = q/C$ and $\Delta V_L = L\,di/dt$, Eq. 37-14 can be written as

$$\mathcal{E} = iR + L\frac{di}{dt} + \frac{q}{C}, \tag{37-25}$$

or, using $i = dq/dt$ along with Eq. 37-1 for the emf,

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = \mathcal{E}_m \sin\omega t. \tag{37-26}$$

This equation is in the same form as that for the forced mechanical oscillator discussed in Section 17-8 (see Eq.

17-42). Making the analogies

$$x \rightarrow q, \quad m \rightarrow L, \quad b \rightarrow R, \quad \text{and} \quad k \rightarrow 1/C,$$

which we also used in Sections 36-5 to 36-7, we can immediately adapt the result given in Eq. 17-43 for the forced, damped mechanical oscillator to the driven, damped (that is, resistive) electromagnetic oscillator:

$$q = -\frac{\mathscr{E}_m}{\omega Z}\cos(\omega t - \phi), \qquad (37\text{-}27)$$

where, as you should show, ωZ is G as defined by Eq. 17-44. Differentiating Eq. 37-27 to find the current, we obtain Eq. 37-1, $i = i_m \sin(\omega t - \phi)$, with $i_m = \mathscr{E}_m/Z$. You should also show that the phase β given by Eq. 17-45 reduces to Eq. 37-18 when we replace the mechanical quantities by their electromagnetic analogues.

Seeking analogies, such as we have done here between mechanical and electromagnetic resonance, is a useful technique that not only provides insight into new phenomena but also saves work in their analysis, because we can adapt mathematical results obtained for one system to the analysis of another. We recognize the common characteristics of the two systems: a sinusoidal driving element; an inertial element, which resists changes in motion (m, which resists changes in v, and L, which resists changes in i); a dissipative element (b and R, each part of terms linear in the rate of change of the coordinate); and a restoring element (k and $1/C$, each part of terms linear in the coordinate). Common features of both solutions are a stable sinusoidal oscillation at the driving frequency after an initial period of rapidly decaying transients, a phase difference between the driver and oscillating coordinate that is independent of the driving amplitude, and resonance at a particular frequency whose value is determined only by the inertial and restoring elements. ■

37-4 POWER IN AC CIRCUITS

In an electrical circuit, energy is supplied by the source of emf, stored by the capacitive and inductive elements, and dissipated in resistive elements. Conservation of energy requires that, at any particular time, the rate at which energy is supplied by the source of emf must equal the rate at which it is stored in the capacitive and inductive elements plus the rate at which it is dissipated in the resistive elements. (We assume ideal capacitive and inductive elements that have no internal resistance.)

Let us consider a resistor as an isolated element (as in Fig. 37-2) in an AC circuit in which the current is given by Eq. 37-2. (We examine the circuit in its steady state, a sufficiently long time after the source of emf has been connected to the circuit.) Just as in a DC circuit, the rate of energy dissipation (Joule heating) in a resistor in an AC circuit is given by

$$P = i^2R = i_m^2 R \sin^2(\omega t - \phi). \qquad (37\text{-}28)$$

The energy dissipated in the resistor fluctuates with time, as does the energy stored in the inductive or capacitive elements. In most cases involving alternating currents, it is of no interest how the power varies during each cycle; the main interest is the *average* power dissipated during any particular cycle. The *average* energy stored in the inductive or capacitive elements remains constant over any complete cycle; in effect, energy is transferred from the source of emf to the resistive elements, where it is dissipated.

For example, the commercial power company supplies an AC source of emf to your home that varies with a frequency of $f = 60$ Hz. You are charged for the *average* power you consume; the power company is not concerned with whether you are operating a purely resistive device, in which the maximum power is dissipated in phase with the source of emf, or a partially capacitive or inductive device such as a motor, in which the current maximum (and therefore the power maximum) might occur out of phase with the emf. If the power company measured your energy use in a time smaller than $\frac{1}{60}$ s, they would notice variations in the rate at which you use energy, but in measuring over a time longer than $\frac{1}{60}$ s only the average rate of energy consumption becomes important.

We write the average power P_{av} by taking the average value of Eq. 37-28 over a time that is large compared with the period of one cycle. For each full cycle, the average value of the $\sin^2$ is $\frac{1}{2}$. If the number of cycles included in the average is large, fractions of a cycle are unimportant and we can obtain the average power by replacing the $\sin^2$ in Eq. 37-28 with the value $\frac{1}{2}$. The average power is then

$$P_{av} = \tfrac{1}{2} i_m^2 R, \qquad (37\text{-}29)$$

which we can also write as

$$P_{av} = (i_m/\sqrt{2})^2 R. \qquad (37\text{-}30)$$

The quantity $i_m/\sqrt{2}$ is equal to the *root-mean-square* (rms) value of the current:

$$i_{rms} = \frac{i_m}{\sqrt{2}}. \qquad (37\text{-}31)$$

It is the result you would obtain if you first squared the current, then took its average (or mean) over a whole number of cycles, and then took the square root. (We defined the rms molecular speed in the same way in Chapter 22.) It is convenient to write the power in terms of rms values, because AC current and voltage meters are designed to report rms values. The common 120 V of household wiring is a rms value; the peak voltage is $\mathscr{E}_m = \sqrt{2}\mathscr{E}_{rms} = \sqrt{2}(120\text{ V}) = 170$ V.

In terms of i_{rms}, Eq. 37-30 can be written

$$P_{av} = i_{rms}^2 R. \qquad (37\text{-}32)$$

Equation 37-32 is similar to the expression $P = i^2R$, which describes the power dissipated in a resistor in a DC circuit. If we replace DC currents and voltages with the rms values of AC currents and voltages, then DC expressions for power dissipation can be used to obtain the average AC power dissipation.

Power in the Series *RLC* Circuit

So far we have been considering the power dissipated in an isolated resistive element in an AC circuit. Let us now consider a full AC circuit from the standpoint of power dissipation. For this purpose we again choose the series *RLC* circuit as an example.

The work dW done by a source of emf $\mathcal{E}$ on a charge dq is given by $dW = \mathcal{E}\, dq$. The power P $(= dW/dt)$ supplied by the source of emf is then $\mathcal{E}\, dq/dt = \mathcal{E}i$, or, using Eqs. 37-1 and 37-2,

$$P = \mathcal{E}i = \mathcal{E}_m i_m \sin \omega t \sin (\omega t - \phi). \qquad (37\text{-}33)$$

We are seldom interested in this instantaneous power, which is usually a rapidly fluctuating function of the time. To find the *average* power, let us first use a trigonometric identity to expand the factor $\sin (\omega t - \phi)$:

$$\begin{aligned} P &= \mathcal{E}_m i_m \sin \omega t (\sin \omega t \cos \phi - \cos \omega t \sin \phi) \\ &= \mathcal{E}_m i_m (\sin^2 \omega t \cos \phi - \sin \omega t \cos \omega t \sin \phi). \qquad (37\text{-}34) \end{aligned}$$

When we now average over a complete cycle, the $\sin^2 \omega t$ term gives the value $\frac{1}{2}$, whereas the $\sin \omega t \cos \omega t$ term gives 0, as you should show (see Exercise 16). The average power is then

$$P_{av} = \tfrac{1}{2}\mathcal{E}_m i_m \cos \phi. \qquad (37\text{-}35)$$

Replacing both $\mathcal{E}_m$ and i_m with their rms values ($\mathcal{E}_{rms} = \mathcal{E}_m/\sqrt{2}$ and $i_{rms} = i_m/\sqrt{2}$), we can write Eq. 37-35 as

$$P_{av} = \mathcal{E}_{rms} i_{rms} \cos \phi. \qquad (37\text{-}36)$$

The quantity $\cos \phi$ in Eq. 37-36 is called the *power factor* of the AC circuit. Let us evaluate the power factor for the series *RLC* circuit. From Eq. 37-18, $\tan \phi = (X_L - X_C)/R$, we can show that

$$\cos \phi = \frac{R}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{R}{Z}. \qquad (37\text{-}37)$$

According to Eq. 37-36, the power delivered to the circuit by the source of emf is maximum when $\cos \phi = 1$, which occurs when the circuit is purely resistive and contains no capacitors or inductors, or at resonance when $X_L = X_C$ so that $Z = R$. In this case the average power is

$$P_{av} = \mathcal{E}_{rms} i_{rms} \quad \text{(resistive load).} \qquad (37\text{-}38)$$

If the load is strongly inductive, as it often is in the case of motors, compressors, and the like, the power delivered to the load can be maximized by increasing the capacitance of the circuit. Power companies often place capacitors throughout their transmission system to bring this about.

SAMPLE PROBLEM 37-5. Consider again the circuit of Fig. 37-1, using the same parameters that we used in Sample Problem 37-3: $R = 160\ \Omega$, $C = 15\ \mu\text{F}$, $L = 230$ mH, $f = 60$ Hz, and $\mathcal{E}_m = 36$ V. Find (*a*) the rms emf, (*b*) the rms current, (*c*) the power factor, and (*d*) the average power dissipated in the resistor.

Solution (*a*)

$$\mathcal{E}_{rms} = \mathcal{E}_m/\sqrt{2} = (36\ \text{V})/\sqrt{2} = 25.5\ \text{V}.$$

(*b*) In Sample Problem 37-3 we found $i_m = 0.196$ A. We then have

$$i_{rms} = i_m/\sqrt{2} = (0.196\ \text{A})/\sqrt{2} = 0.139\ \text{A}.$$

(*c*) In Sample Problem 37-3 we found that the phase constant ϕ was $-29.4°$. Thus

$$\text{power factor} = \cos(-29.4°) = 0.871.$$

(*d*) From Eq. 37-32 we have

$$P_{av} = i_{rms}^2 R = (0.139\ \text{A})^2(160\ \Omega) = 3.1\ \text{W}.$$

Alternatively, Eq. 37-36 yields

$$\begin{aligned} P_{av} &= \mathcal{E}_{rms} i_{rms} \cos \phi \\ &= (25.5\ \text{V})(0.139\ \text{A})(0.871) = 3.1\ \text{W}, \end{aligned}$$

in full agreement. That is, the average power dissipated in the resistor equals the average power supplied by the emf. In effect, energy is transferred from the emf to the resistive load, where it is dissipated. Note that, to get agreement of these results to two significant figures, we had to use three significant figures for the currents and voltages. Aside from such numerical rounding errors, Eqs. 37-32 and 37-36 give identical results.

37-5 THE TRANSFORMER (Optional)

In DC circuits the power dissipation in a resistive load is given by Eq. 31-21 ($P_R = i\,\Delta V_R = i^2R$). For a given power requirement, we have our choice of a relatively large current i and a relatively small potential difference ΔV_R or exactly the reverse, provided that their product remains constant. In the same way, for purely resistive AC circuits (in which the power factor, $\cos \phi$ in Eq. 37-36, is equal to 1), the average power dissipation is given by Eq. 37-38 ($P_{av} = i_{rms}\mathcal{E}_{rms}$) and we have the same choice as to the relative values of i_{rms} and $\mathcal{E}_{rms}$.

In electric power distribution systems it is desirable, both for reasons of safety and the efficient design of equipment, to have relatively low voltages at both the generating end (the electric power plant) and the receiving end (the home or factory). For example, no one wants an electric toaster or a child's electric train to operate at, say, 10 kV.

On the other hand, in the transmission of electric energy from the generating plant to the consumer, we want the *lowest* practical current (and thus the *largest* practical potential difference) so as to minimize the i^2R energy dissipation in the transmission line. Values such as $\mathcal{E}_{rms} = 350$ kV are typical. Thus there is a fundamental mismatch between the requirements for efficient transmission on the one hand and efficient and safe generation and consumption on the other hand.

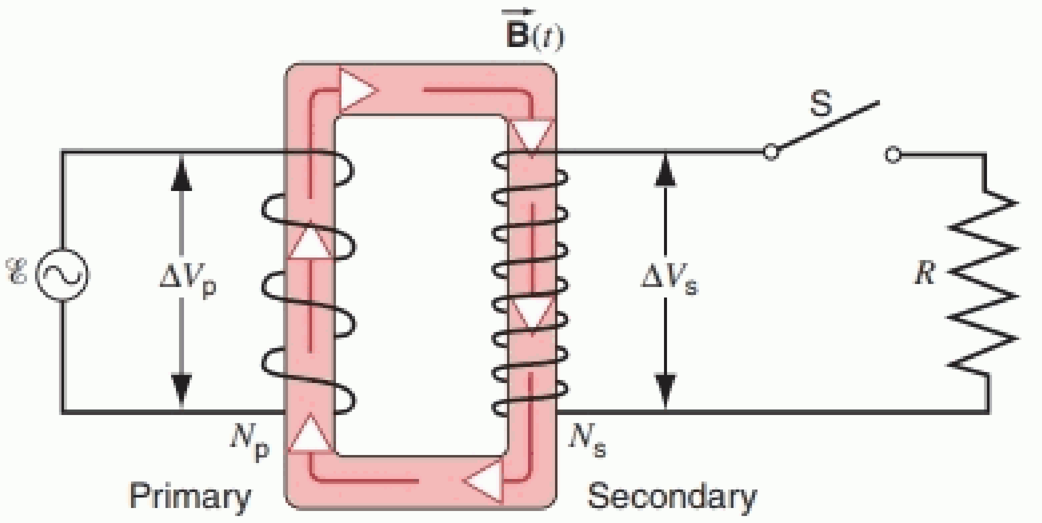


FIGURE 37-6. An ideal transformer, showing two coils wound on an iron core.

To overcome this problem, we need a device that can, as design considerations require, raise (or lower) the potential difference in a circuit, keeping the product $i_{rms}\mathscr{E}_{rms}$ essentially constant. The alternating current *transformer* of Fig. 37-6 is such a device. Operating on the basis of Faraday's law of induction, the transformer has no direct current counterpart of equivalent simplicity, which is why DC distribution systems, strongly advocated by Edison, were replaced by AC systems, strongly advocated by Tesla and others.*

In Fig. 37-6, two coils are shown wound around an iron core. The *primary* winding, of N_p turns, is connected to an alternating current generator whose emf $\mathscr{E}$ is given by $\mathscr{E} = \mathscr{E}_m \sin \omega t$. The *secondary* winding, of N_s turns, is an open circuit as long as switch S is open, which we assume for the present. Thus there is no current in the secondary winding. We assume further that we can neglect all dissipative elements, such as the resistances of the primary and secondary windings. Actually, well-designed, high-capacity transformers can have energy losses as low as 1%, so our assumption of an ideal transformer is not unreasonable.

For the above conditions the primary winding is a pure inductance, as in Fig. 37-3*a*. The (very small) primary current, called the *magnetizing current* $i_{mag}(t)$, lags the primary potential difference $\Delta V_p(t)$ by 90°; the power factor ($=\cos\phi$ in Eq. 37-36) is zero, so no power is delivered from the generator to the transformer.

However, the small alternating primary current $i_{mag}(t)$ induces an alternating magnetic flux $\Phi_B(t)$ in the iron core, and we assume that all this flux links the turns of the secondary windings. (That is, we assume that all the magnetic field lines form closed loops within the iron core and none "escape" into the surroundings.) From Faraday's law of induction the emf *per turn* $\mathscr{E}_T$ (equal to $-d\Phi_B/dt$) is the same for both the primary and secondary windings, because the primary and secondary fluxes are equal. On a rms basis, we can write

$$\left(\frac{d\Phi_B}{dt}\right)_{\text{primary}} = \left(\frac{d\Phi_B}{dt}\right)_{\text{secondary}} \tag{37-39}$$

or

$$(\mathscr{E}_T)_{\text{rms, primary}} = (\mathscr{E}_T)_{\text{rms, secondary}}. \tag{37-40}$$

For each winding, the emf per turn equals the potential difference divided by the number of turns in the winding; Eq. 37-40 can then be written

$$\frac{\Delta V_p}{N_p} = \frac{\Delta V_s}{N_s}. \tag{37-41}$$

Here ΔV_p and ΔV_s refer to rms quantities. Solving for ΔV_s, we obtain

$$\Delta V_s = \Delta V_p(N_s/N_p). \tag{37-42}$$

If $N_s > N_p$ (in which case, $\Delta V_s > \Delta V_p$), we speak of a *step-up transformer;* if $N_s < N_p$, we speak of a *step-down transformer.*

In all of the above we have assumed an open circuit secondary so that no power is transmitted through the transformer. If we now close switch S in Fig. 37-6, however, we have a more practical situation in which the secondary winding is connected with a resistive load *R*. In the general case, the load would also contain inductive and capacitive elements, but we confine ourselves to this special case of a purely resistive load.

Several things happen when we close switch S. (1) A rms current i_s appears in the secondary circuit, with a corresponding average power dissipation i_s^2R in the resistive load. (2) The alternating secondary current induces its own alternating magnetic flux in the iron core, and this flux induces (from Faraday's law and Lenz' law) an opposing emf in the primary windings. (3) ΔV_p, however, cannot change in response to this opposing emf because it must always equal the emf that is provided by the generator; closing switch S cannot change this fact. (4) To ensure this, a new alternating current i_p must appear in the primary circuit, its magnitude and phase constant being exactly that needed to cancel the opposing emf generated in the primary windings by i_s.

Rather than analyze the above rather complex process in detail, we take advantage of the overall view provided by the conservation-of-energy principle. For an ideal transformer with a resistive load, this tells us that

$$i_p\,\Delta V_p = i_s\,\Delta V_s. \tag{37-43}$$

Because Eq. 37-42 holds whether or not the switch S of Fig. 37-6 is closed, we then have

$$i_s = i_p(N_p/N_s) \tag{37-44}$$

as the transformation relation for currents.

Finally, knowing that $i_s = \Delta V_s/R$, we can use Eqs. 37-42 and 37-44 to obtain

$$i_p = \frac{\Delta V_p}{(N_p/N_s)^2 R}, \tag{37-45}$$

which tells us that, from the point of view of the primary circuit, the equivalent resistance of the load is not *R* but

$$R_{eq} = (N_p/N_s)^2 R. \tag{37-46}$$

---

* See "The Transformer," by John W. Coltman, *Scientific American,* January 1988, p. 86.

Equation 37-46 suggests still another function for the transformer. We have seen that, for maximum transfer of energy from a seat of emf to a resistive load, the resistance of the generator and the resistance of the load must be equal. (See Exercise 14 of Chapter 31.) The same relation holds for AC circuits except that the *impedance* (rather than the resistance) of the generator must be matched to that of the load. It often happens—as when we wish to connect a speaker to an amplifier—that this condition is far from met, the amplifier being of high impedance and the speaker of low impedance. We can match the impedances of the two devices by coupling them through a transformer with a suitable turns ratio.

SAMPLE PROBLEM 37-6. A transformer on a utility pole operates at $\Delta V_p = 8.5$ kV on the primary side and supplies electric energy to a number of nearby houses at $\Delta V_s = 120$ V, both quantities being rms values. The rate of average energy consumption in the houses served by the transformer at a given time is 78 kW. Assume an ideal transformer, a resistive load, and a power factor of unity. (*a*) What is the turns ratio N_p/N_s of this step-down transformer? (*b*) What are the rms currents in the primary and secondary windings of the transformer? (*c*) What is the equivalent resistive load in the secondary circuit? (*d*) What is the equivalent resistive load in the primary circuit?

Solution (*a*) From Eq. 37-42 we have

$$\frac{N_p}{N_s} = \frac{\Delta V_p}{\Delta V_s} = \frac{8.5 \times 10^3\ \text{V}}{120\ \text{V}} = 70.8.$$

(*b*) From Eq. 37-38,

$$i_p = \frac{P_{\text{av}}}{\Delta V_p} = \frac{78 \times 10^3\ \text{W}}{8.5 \times 10^3\ \text{V}} = 9.18\ \text{A}$$

and

$$i_s = \frac{P_{\text{av}}}{\Delta V_s} = \frac{78 \times 10^3\ \text{W}}{120\ \text{V}} = 650\ \text{A}.$$

(*c*) In the secondary circuit,

$$R_s = \frac{\Delta V_s}{i_s} = \frac{120\ \text{V}}{650\ \text{A}} = 0.185\ \Omega.$$

(*d*) Here we have

$$R_p = \frac{\Delta V_p}{i_p} = \frac{8.5 \times 10^3\ \text{V}}{9.18\ \text{A}} = 930\ \Omega.$$

We can verify this from Eq. 37-46, which we write as

$$R_p = (N_p/N_s)^2 R_s = (70.8)^2(0.185\ \Omega) = 930\ \Omega.$$

MULTIPLE CHOICE

37-1 Alternating Currents

37-2 Three Separate Elements

1. Which of the following quantities *increases* with increasing frequency?

(A) R (B) L (C) C (D) X_L (E) X_C

2. Which circuit element would make the best series filter to prevent high-frequency signals from passing through a bass speaker?

(A) A capacitor (B) An inductor
(C) A resistor (D) A transformer

3. AC current flows through a capacitor from right to left as shown in Fig. 37-7. The current on the right is i_1, and the current on the left is i_2. As the AC frequency is decreased, what happens to i_1 and i_2?

(A) i_1 remains constant, and i_2 decreases.
(B) i_1 decreases, but i_2 decreases at a faster rate.
(C) The difference $i_1 - i_2$ increases.
(D) The difference $i_1 - i_2$ remains fixed.

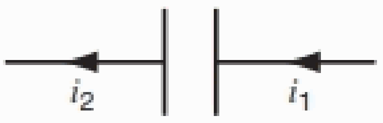


FIGURE 37-7. Multiple-choice question 3.

4. Which type of material, if inserted into an inductor, will cause the largest increase in inductive reactance?

(A) Dielectric (B) Diamagnetic
(C) Paramagnetic (D) Ferromagnetic

37-3 The Single-Loop *RLC* Circuit

5. A series *RLC* circuit powered by a source $\mathcal{E} = \mathcal{E}_m \sin \omega t$ is operating at a frequency that is less than the resonant frequency. A dielectric slab is inserted between the plates of the capacitor. How does i_m change?

(A) i_m increases. (B) i_m decreases.
(C) i_m remains the same.
(D) i_m could increase or decrease, but it does not remain the same.
(E) There is not enough information to answer the question.

6. A student constructs a series *RLC* circuit. While operating the circuit at a frequency f she uses an AC voltmeter and measures the potential difference across each device as $(\Delta V_R)_{\max} = 8.8$ V, $(\Delta V_L)_{\max} = 2.6$ V, and $(\Delta V_C)_{\max} = 7.4$ V.

(*a*) The circuit is constructed so that the inductor is next to the capacitor. What result should the student expect for a measurement of the combined potential difference $(\Delta V_L + \Delta V_C)_{\max}$ across the inductor and capacitor?

(A) 10.0 V (B) 7.8 V (C) 7.4 V (D) 4.8 V

(*b*) What result should the student expect for a measurement of the amplitude $\mathcal{E}_m$ of the potential difference across the power supply?

(A) 18.8 V (B) 13.6 V (C) 10.0 V (D) 4.0 V

7. A student constructs a series *RLC* circuit. While operating the circuit at a frequency f, he uses an AC voltmeter and measures the potential difference across each device as $(\Delta V_R)_{\rm max} = 4.8$ V, $(\Delta V_L)_{\rm max} = 29$ V, and $(\Delta V_C)_{\rm max} = 20$ V.

(*a*) How should the frequency of this circuit be changed to *increase* the current $i_{\rm m}$ through the circuit?

(A) Increase f. (B) Decrease f.
(C) The current is already at a maximum.
(D) There is not enough information to answer the question.

(*b*) What will happen to the value of $(\Delta V_L)_{\rm max}$ if the frequency is adjusted to increase the current through the circuit?

(A) $(\Delta V_L)_{\rm max}$ will increase.
(B) $(\Delta V_L)_{\rm max}$ will decrease.
(C) $(\Delta V_L)_{\rm max}$ will remain the same regardless of any changes to f.
(D) The current is already at a maximum.
(E) There is not enough information to answer the question.

8. Which of the following can never occur in a series *RLC* circuit?

(A) $(\Delta V_R)_{\rm max} > \mathcal{E}_{\rm m}$ (B) $(\Delta V_L)_{\rm max} > \mathcal{E}_{\rm m}$
(C) $(\Delta V_C)_{\rm max} > \mathcal{E}_{\rm m}$ (D) $(\Delta V_L)_{\rm max} > (\Delta V_C)_{\rm max}$

37-4 Power in AC Circuits

9. What happens to the "missing" power in Eq. 37-35 when $\phi > 0$?

(A) It is lost through the capacitor.
(B) It is lost through the inductor.
(C) It is lost through the resistor.
(D) There is no "missing" power; Eq. 37-35 accounts for everything.

37-5 The Transformer

10. What is the effect on the equivalent resistance of a transformer when doubling the number of turns on the primary coil while halving the number of turns on the secondary coil?

(A) The equivalent resistance increases 64 fold.
(B) The equivalent resistance increases 16 fold.
(C) The equivalent resistance increases 4 fold.
(D) The equivalent resistance decreases to 1/4 of the previous value.

QUESTIONS

1. In the relation $\omega = 2\pi f$, when using SI units we measure ω in radians per second and f in hertz or cycles per second. The radian is a measure of angle. What connection do angles have with alternating current?

2. If the output of an AC generator such as that in Fig. 34-13 is connected to an *RLC* circuit such as that of Fig. 37-1, what is the ultimate source of the energy dissipated in the resistor?

3. Why would power distribution systems be less effective without alternating current?

4. In the circuit of Fig. 37-1, why is it safe to assume that (*a*) the alternating current of Eq. 37-2 has the same angular frequency ω as the alternating emf of Eq. 37-1, and (*b*) that the phase angle ϕ in Eq. 37-2 does not vary with time? What would happen if either of these (true) statements were false?

5. How does a phasor differ from a vector? We know, for example, that emfs, potential differences, and currents are not vectors. How then can we justify constructions such as Fig. 37-5?

6. In the purely resistive circuit element of Fig. 37-2, does the maximum value $i_{\rm m}$ of the alternating current vary with the angular frequency of the applied emf?

7. Would any of the discussion of Section 37-3 be invalid if the phasor diagrams were to rotate clockwise, rather than counterclockwise as we assumed?

8. Suppose that, in a series *RLC* circuit, the frequency of the applied voltage is changed continuously from a very low value to a very high value. How does the phase constant change?

9. Could the alternating current resistance of a device depend on the frequency?

10. From the analysis of an *RLC* circuit we can determine the behavior of an *RL* circuit (no capacitor) by putting $C = \infty$, whereas we put $L = 0$ to determine the behavior of an *RC* circuit (no inductor). Explain this difference.

11. During World War II, an AC generator was located a mile or so from the laboratory building it served. A technician increased the speed of the generator to compensate for what he called "the loss of frequency along the transmission line" connecting the generator with the laboratory building. Comment on this procedure.

12. As the speed of the blades of a rotating fan is increased from zero, a series of stationary patterns can be observed when the blades are illuminated by light from an AC source. The effect is more pronounced when a fluorescent tube or neon lamp is used than it is with a tungsten filament lamp. Explain these observations.

13. Assume that in Fig. 37-1 we let $\omega \rightarrow 0$. Does Eq. 37-19 approach an expected value? What is this value? Discuss.

14. Discuss in your own words what it means to say that an alternating current "leads" or "lags" an alternating emf.

15. If, as we stated in Section 37-3, a given circuit is "more inductive than capacitive"—that is, that $X_L > X_C$—(*a*) does this mean, for a fixed angular frequency, that L is relatively "large" and C is relatively "small," or L and C are both relatively "large"? (*b*) For fixed values of L and C does this mean that ω is relatively "large" or relatively "small"?

16. How could you determine, in a series *RLC* circuit, whether the circuit frequency is above or below resonance?

17. Criticize this statement: "If $X_L > X_C$, then we must have $L > 1/C$."

18. How, if at all, must Kirchhoff's rules (the loop and junction rules) for direct current circuits be modified when applied to AC circuits?

19. Do the loop rule and the junction rule apply to multiloop AC circuits as well as to multiloop DC circuits?

20. In Sample Problem 37-5, what would be the effect on $P_{\rm av}$ if you increased (*a*) R, (*b*) C, and (*c*) L? (*d*) How would ϕ in Eq. 37-36 change in these three cases?

21. If $R = 0$ in the circuit of Fig. 37-1, there can be no power dissipation in the circuit. However, an alternating emf and an alternating current are still present. Discuss the energy flow in the circuit under these conditions.

22. Is there an rms power of an AC circuit?

23. Do commercial power station engineers like to have a low power factor or a high one, or does it make any difference to them? Between what values can the power factor range? What determines the power factor; is it characteristic of the generator, of the transmission line, of the circuit to which the transmission line is connected, or some combination of these?

24. Can the instantaneous power delivered by a source of alternating current ever be negative? Can the power factor ever be negative? If so, explain the meaning of these negative values.

25. In a series *RLC* circuit the emf is leading the current for a particular frequency of operation. You now lower the frequency slightly. Does the total impedance of the circuit increase, decrease, or stay the same?

26. If you know the power factor ($= \cos\phi$ in Eq. 37-36) for a given *RLC* circuit, can you tell whether or not the applied alternating emf is leading or lagging the current? If so, how? If not, why not?

27. What is the permissible range of values of the phase angle ϕ in Eq. 37-2? Of the power factor in Eq. 37-36?

28. Why is it useful to use the rms notation for alternating currents and voltages?

29. You want to reduce your electric bill. Do you hope for a small or a large power factor or does it make any difference? If it does, is there anything you can do about it? Discuss.

30. In Eq. 37-36, is ϕ the phase angle between $\mathcal{E}(t)$ and $i(t)$ or between $\mathcal{E}_{\text{rms}}$ and i_{rms}? Explain.

31. A doorbell transformer is designed for a primary rms input of 120 V and a secondary rms output of 6 V. What would happen if the primary and secondary connections were accidentally interchanged during installation? Would you have to wait for someone to push the doorbell to find out? Discuss.

32. You are given a transformer enclosed in a wooden box, its primary and secondary terminals being available at two opposite faces of the box. How could you find its turns ratio without opening the box?

33. In the transformer of Fig. 37-6, with the secondary an open circuit, what is the phase relationship between (*a*) the applied emf and the primary current, (*b*) the applied emf and the magnetic field in the transformer core, and (*c*) the primary current and the magnetic field in the transformer core?

34. What are some applications of a step-up transformer? Of a step-down transformer?

35. What determines which winding of a transformer is the primary and which the secondary? Can a transformer have a single primary and two secondaries? A single secondary and two primaries?

36. Instead of the 120-V, 60-Hz current typical of the United States, Europe uses 240-V, 50-Hz alternating currents. While on vacation in Europe, you would like to use some of your American appliances, such as a clock, an electric razor, and a hair dryer. Can you do so simply by plugging in a 2:1 step-up transformer? Explain why this apparently simple step may or may not suffice.

EXERCISES

37-1 Alternating Currents

37-2 Three Separate Elements

1. Let Eq. 37-1 describe the effective emf available at an ordinary 60-Hz AC outlet. To what angular frequency ω does this correspond? How does the utility company establish this frequency?

2. A 45.2-mH inductor has a reactance of 1.28 kΩ. (*a*) Find the frequency. (*b*) What is the capacitance of a capacitor with the same reactance at that frequency? (*c*) If the frequency is doubled, what are the reactances of the inductor and capacitor?

3. (*a*) At what angular frequency would a 6.23-mH inductor and a 11.4-μF capacitor have the same reactance? (*b*) What would this reactance be? (*c*) Show that this frequency would be equal to the natural frequency of free *LC* oscillations.

4. The output of an AC generator is $\mathcal{E} = \mathcal{E}_{\text{m}} \sin \omega t$, with $\mathcal{E}_{\text{m}} =$ 25.0 V and $\omega = 377$ rad/s. It is connected to a 12.7-H inductor. (*a*) What is the maximum value of the current? (*b*) When the current is a maximum, what is the emf of the generator? (*c*) When the emf of the generator is -13.8 V and increasing in magnitude, what is the current? (*d*) For the conditions of part (*c*), is the generator supplying energy to or taking energy from the rest of the circuit?

5. The AC generator of Exercise 4 is connected to a 4.1 μF capacitor. (*a*) What is the maximum value of the current? (*b*) When the current is a maximum, what is the emf of the generator? (*c*) When the emf of the generator is -13.8 V and increasing in magnitude, what is the current? (*d*) For the conditions of part (*c*), is the generator supplying energy to or taking energy from the rest of the circuit?

37-3 The Single-Loop *RLC* Circuit

6. A coil of inductance 88.3 mH and unknown resistance and a 937-nF capacitor are connected in series with an oscillator of frequency 941 Hz. The phase angle ϕ between the applied emf and current is 75.0°. Find the resistance of the coil.

7. Redraw (roughly) Figs. 37-5*b* and 37-5*c* for the cases of $X_C > X_L$ and $X_C = X_L$.

8. (*a*) Recalculate all the quantities asked for in Sample Problem 37-3 for $C = 70\ \mu$F, the other parameters in that sample problem remaining unchanged. (*b*) Draw to scale a phasor diagram like that of Fig. 37-5*c* for this new situation and compare the two diagrams closely.

9. Consider the resonance curves of Fig. 36-13. (*a*) Show that for frequencies above resonance the circuit is predominantly inductive and for frequencies below resonance it is predominantly capacitive. (*b*) How does the circuit behave at resonance? (*c*) Sketch a phasor diagram like that of Fig. 37-5*c* for conditions at a frequency higher than resonance, at resonance, and lower than resonance.

10. Verify mathematically that the following geometrical construction correctly gives both the impedance Z and the phase constant ϕ. Referring to Fig. 37-8, (1) draw an arrow in the $+y$ direction of magnitude X_C, (2) draw an arrow in the $-y$ direction of magnitude X_L, and (3) draw an arrow of magnitude R in the $+x$ direction. Then the magnitude of the "resultant" of these arrows is Z and the angle (measured below the $+x$ axis) of this resultant is ϕ.

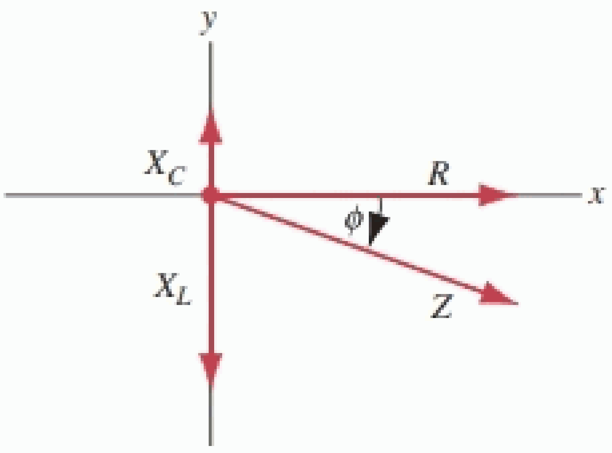


FIGURE 37-8. Exercise 10.

11. Can the amplitude of the voltage across an inductor be greater than the amplitude of the generator emf in an *RLC* circuit? Consider a circuit with $\mathcal{E}_m = 10$ V, $R = 9.6\ \Omega$, $L = 1.2$ H, and $C = 1.3\ \mu$F. Find the amplitude of the voltage across the inductor at resonance.

12. Use Eq. 37-18 to obtain relationships for $\sin\phi$ and $\cos\phi$ in terms of R, X_L, and X_C. Then substitute those expressions into Eq. 37-16 to obtain Eq. 37-17.

13. When the generator emf in Sample Problem 37-3 is a maximum, what is the voltage across (*a*) the generator, (*b*) the resistor, (*c*) the capacitor, and (*d*) the inductor? (*e*) By summing these with appropriate signs, verify that the loop rule is satisfied.

14. A certain *RLC* combination, R_1, L_1, C_1, has a resonant frequency that is the same as that of a different combination, R_2, L_2, C_2. You now connect the two combinations in series. Show that this new circuit has the same resonant frequency as the separate individual circuits.

15. For a certain *RLC* circuit the maximum generator emf is 125 V and the maximum current is 3.20 A. If the current leads the generator emf by 56.3°, (*a*) what is the impedance and (*b*) what is the resistance of the circuit? (*c*) Is the circuit predominantly capacitive or inductive?

37-4 Power in AC Circuits

16. Show that $[\sin^2 \omega t]_{av} = \frac{1}{2}$ and $[\sin \omega t \cos \omega t]_{av} = 0$, where the averages are taken over one or more complete cycles.

17. An electric motor connected to a 120-V, 60-Hz power outlet does mechanical work at the rate of 0.10 hp (1 hp = 746 W). If it draws an rms current of 650 mA, what is its resistance, in terms of power transfer? Would this be the same as the resistance of its coils, as measured with an ohmmeter with the motor disconnected from the power outlet?

18. Show that the average power delivered to an *RLC* circuit can also be written

$$P_{av} = \mathcal{E}_{rms}^2 R/Z^2.$$

Show that this expression gives reasonable results for a purely resistive circuit, for an *RLC* circuit at resonance, for a purely capacitive circuit, and for a purely inductive circuit.

19. Calculate the average power dissipated in Sample Problem 37-3 assuming (*a*) that the inductor is removed from the circuit and (*b*) that the capacitor is removed.

20. An air conditioner connected to a 120-V, rms AC line is equivalent to a 12.2-Ω resistance and a 2.30-Ω inductive reactance in series. (*a*) Calculate the impedance of the air conditioner. (*b*) Find the average power supplied to the appliance. (*c*) What is the value of the rms current?

21. A high-impedance AC voltmeter is connected in turn across the inductor, the capacitor, and the resistor in a series circuit having an AC source of 100 V (rms) and gives the same reading in volts in each case. What is this reading?

22. Consider the FM antenna circuit shown in Fig. 37-9, with $L = 8.22\ \mu$H, $C = 0.270$ pF, and $R = 74.7\ \Omega$. The radio signal induces an alternating emf in the antenna with $\mathcal{E}_{rms} = 9.13\ \mu$V. Find (*a*) the frequency of the incoming waves for which the antenna is "in tune," (*b*) the rms current in the antenna, and (*c*) the rms potential difference across the capacitor.

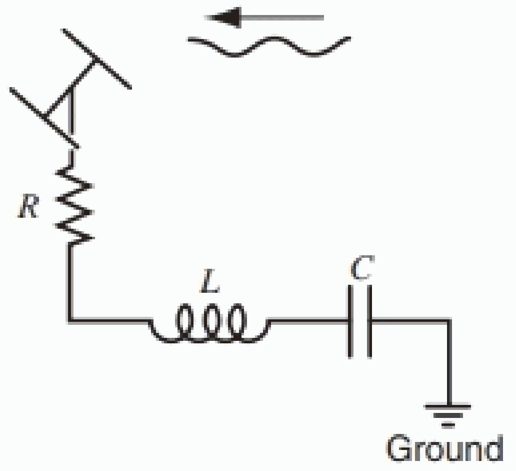


FIGURE 37-9. Exercise 22.

23. In Fig. 37-10, show that the power dissipated in the resistor R is a maximum when $R = r$, in which r is the internal resistance of the AC generator. In the text we have tacitly assumed, up to this point, that $r = 0$. Compare with the DC situation.

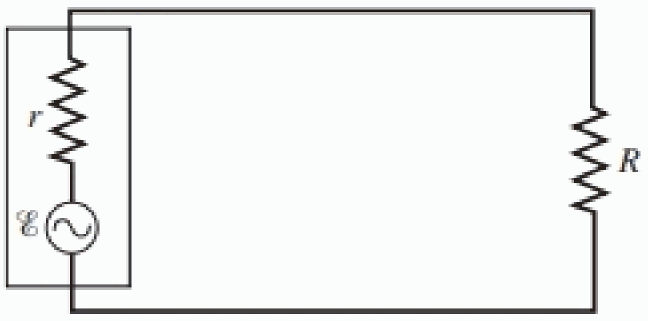


FIGURE 37-10. Exercises 23 and 32.

24. In an *RLC* circuit such as that of Fig. 37-1, assume that $R = 5.0\ \Omega$, $L = 60$ mH, $f = 60$ Hz, and $\mathcal{E}_m = 30$ V. For what values of the capacitance would the average power dissipated in the resistor be (*a*) a maximum and (*b*) a minimum? (*c*) What are these maximum and minimum powers? (*d*) What are the corresponding phase angles? (*e*) What are the corresponding power factors?

25. In Fig. 37-11, $R = 15.0\ \Omega$, $C = 4.72\ \mu$F, and $L = 25.3$ mH. The generator provides a sinusoidal voltage of 75.0 V (rms) and frequency $f = 550$ Hz. (*a*) Calculate the rms current amplitude. (*b*) Find the rms voltages ΔV_{ab}, ΔV_{bc}, ΔV_{cd}, ΔV_{bd}, ΔV_{ad}. (*c*) What average power is dissipated by each of the three circuit elements?

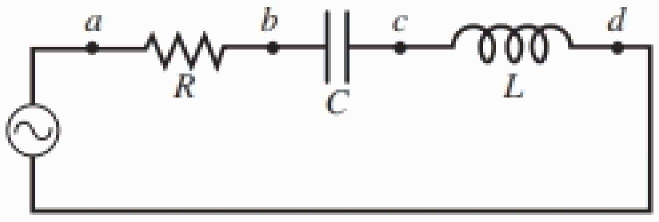


FIGURE 37-11. Exercise 25.

26. For an *RLC* circuit, show that in one cycle with period T (*a*) the energy stored in the capacitor does not change, (*b*) the energy stored in the inductor does not change, (*c*) the generator supplies energy $(\frac{1}{2}T)\mathcal{E}_{\mathrm{m}} i_{\mathrm{m}} \cos\phi$; and (*d*) the resistor dissipates energy $(\frac{1}{2}T)Ri_{\mathrm{m}}^2$. (*e*) Show that the quantities found in (*c*) and (*d*) are equal.

37-5 The Transformer

27. A generator supplies 150 V (rms) to the primary coil of a transformer of 65 turns. If the secondary coil has 780 turns, what is the secondary voltage?

28. A transformer has 500 primary turns and 10 secondary turns. (*a*) If ΔV_{p} for the primary is 120 V (rms), what is ΔV_{s} for the secondary, assumed to be an open circuit? (*b*) If the secondary is now connected to a resistive load of 15 Ω, what are the currents in the primary and secondary windings?

29. Figure 37-12 shows an "autotransformer." It consists of a single coil (with an iron core). Three "taps" are provided. Between taps T_1 and T_2 there are 200 turns and between taps T_2 and T_3 there are 800 turns. Any two taps can be considered the "primary terminals" and any two taps can be considered the "secondary terminals." List all the ratios by which the primary voltage may be changed to a secondary voltage.

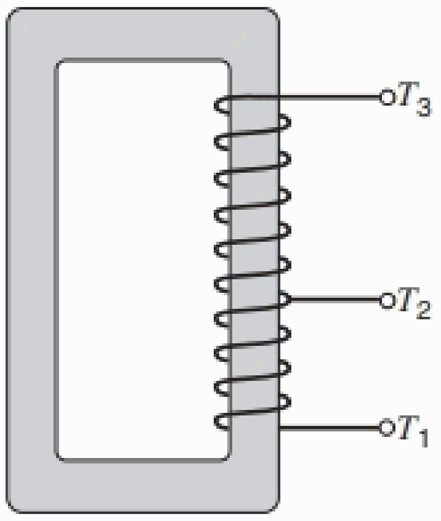


FIGURE 37-12. Exercise 29.

30. A farmer runs a water pump at 3.8 A rms. The connecting line from the transformer is 1.2 km long and consists of two copper wires, each 1.8 mm in diameter. The temperature is 5.4°C. How much power is lost in transmission through the line?

31. An electrical engineer designs an ideal transformer to run an x-ray machine at a peak potential of 74 kV and 270-mA rms current. The transformer operates from a 220-V rms power supply. However, resistance in the wires connecting the power supply to the transformer was ignored. Upon installation, it is realized that the supply wires have a resistance of 0.62 Ω. By how much must the supply voltage be increased in order to maintain the same operating parameters at the transformer?

32. In Fig. 37-10, let the rectangular box on the left represent the (high-impedance) output of an audio amplifier, with $r = 1000\ \Omega$. Let $R = 10\ \Omega$ represent the (low-impedance) coil of a loudspeaker. We learned that a transformer can be used to "transform" resistances, making them behave electrically as if they were larger or smaller than they actually are. Sketch the primary and secondary coils of a transformer to be introduced between the "amplifier" and the "speaker" in Fig. 37-10 to "match the impedances." What must be the turns ratio?

PROBLEMS

1. The output of an AC generator is given by $\mathcal{E} = \mathcal{E}_{\mathrm{m}} \sin(\omega t - \pi/4)$, where $\mathcal{E}_{\mathrm{m}} = 31.4$ V and $\omega = 350$ rad/s. The current is given by $i(t) = i_{\mathrm{m}} \sin(\omega t - 3\pi/4)$, where $i_{\mathrm{m}} = 622$ mA. (*a*) At what time, after $t = 0$, does the generator emf first reach a maximum? (*b*) At what time, after $t = 0$, does the current first reach a maximum? (*c*) The circuit contains a single element other than the generator. Is it a capacitor, an inductor, or a resistor? Justify your answer. (*d*) What is the value of the capacitance, inductance, or resistance, as the case may be?

2. Repeat Problem 1 except that now $i = i_{\mathrm{m}} \sin(\omega t + \pi/4)$.

3. In a certain *RLC* circuit, operating at 60 Hz, the maximum voltage across the inductor is twice the maximum voltage across the resistor, while the maximum voltage across the capacitor is the same as the maximum voltage across the resistor. (*a*) By what phase angle does the current lag the generator emf? (*b*) If the maximum generator emf is 34.4 V, what should be the resistance of the circuit to obtain a maximum current of 320 mA?

4. The AC generator in Fig. 37-13 supplies 170 V (max) at 60 Hz. With the switch open as in the diagram, the resulting current leads the generator emf by 20°. With the switch in position 1 the current lags the generator emf by 10°. When the switch is in position 2 the maximum current is 2.82 A. Find the values of *R*, *L*, and *C*.

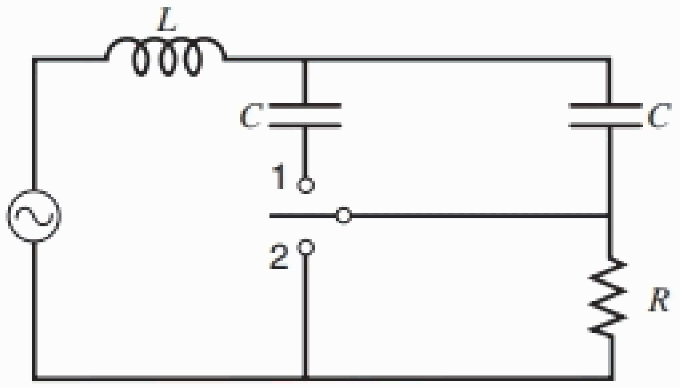


FIGURE 37-13. Problem 4.

5. A three-phase generator G produces electrical power that is transmitted by means of three wires as shown in Fig. 37-14. The potentials (relative to a common reference level) of these wires are $V_1 = V_m \sin \omega t$, $V_2 = V_m \sin(\omega t - 120°)$, and $V_3 = V_m \sin(\omega t - 240°)$. Some industrial equipment (for example, motors) has three terminals and is designed to be connected directly to these three wires. To use a more conventional two-terminal device (for example, a lightbulb), one connects it to any two of the three wires. Show that the potential difference between any two of the wires (*a*) oscillates sinusoidally with angular frequency ω and (*b*) has amplitude $V_m\sqrt{3}$.

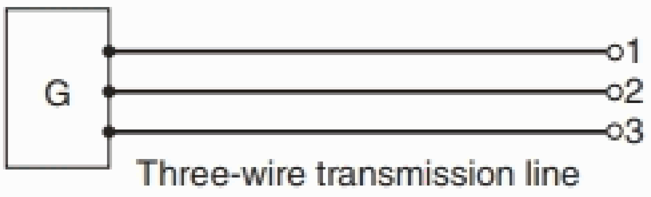


FIGURE 37-14. Problem 5.

6. Figure 37-15 shows an AC generator connected to a "black box" through a pair of terminals. The box contains an *RLC* circuit, possibly even a multiloop circuit, whose elements and arrangements we do not know. Measurements outside the box reveal that $\mathscr{E}(t) = (75\text{ V}) \sin \omega t$ and $i(t) = (1.2\text{ A}) \sin(\omega t + 42°)$. (*a*) What is the power factor? (*b*) Does the current lead or lag the emf? (*c*) Is the circuit in the box largely inductive or largely capacitive in nature? (*d*) Is the circuit in the box in resonance? (*e*) Must there be a capacitor in the box? An inductor? A resistor? (*f*) What average power is delivered to the box by the generator? (*g*) Why do you not need to know the angular frequency ω to answer all these questions?

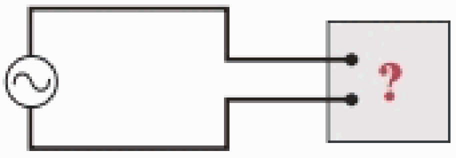


FIGURE 37-15. Problem 6.

7. An *RLC* circuit has $R = 5.12\ \Omega$, $C = 19.3\ \mu\text{F}$, $L = 988$ mH, and $\mathscr{E}_m = 31.3$ V. (*a*) At what angular frequency ω will the current have its maximum value, as in the resonance curves of Fig. 36-13? (*b*) What is this maximum value? (*c*) At what two angular frequencies ω_1 and ω_2 will the current amplitude have one-half of this maximum value? (*d*) Find the fractional width $[= (\omega_1 - \omega_2)/\omega]$ of the resonance curve.

8. (*a*) Show that the fractional width of the resonance curves of Fig. 36-13 is given, to a close approximation, by

$$\frac{\Delta\omega}{\omega} = \frac{\sqrt{3}R}{\omega L},$$

in which ω is the resonant frequency and $\Delta\omega$ is the width of the resonance peak at $i = \frac{1}{2}i_m$. Note (see Problem 15 of Chapter 36) that this expression may be written as $\sqrt{3}/Q$, which shows clearly that a "high-Q" circuit has a sharp resonance peak—that is, a small value of $\Delta\omega/\omega$. (*b*) Use this result to check part (*d*) of Problem 7.

9. In an *RLC* circuit, $R = 16.0\ \Omega$, $C = 31.2\ \mu\text{F}$, $L = 9.20$ mH, $\mathscr{E} = \mathscr{E}_m \sin \omega t$ with $\mathscr{E}_m = 45.0$ V, and $\omega = 3000$ rad/s. For time $t = 0.442$ ms, find (*a*) the rate at which energy is being supplied by the generator, (*b*) the rate at which energy is being stored in the capacitor, (*c*) the rate at which energy is being stored in the inductor, and (*d*) the rate at which energy is being dissipated in the resistor. (*e*) What is the meaning of a negative result for any of parts (*a*), (*b*), and (*c*)? (*f*) Show that the results of parts (*b*), (*c*), and (*d*) sum to the result of part (*a*).

10. In Fig. 37-6, show that $i_p(t)$ in the primary circuit remains unchanged if a resistance R' $[= R(N_p/N_s)^2]$ is connected directly across the generator, the transformer and the secondary circuit being removed. That is,

$$i_p(t) = \frac{\mathscr{E}(t)}{R'}.$$

In this sense we see that a transformer "transforms" not only potential differences and currents but also resistances. In the more general case, in which the secondary load in Fig. 37-6 contains capacitive and inductive elements as well as resistive, we say that a transformer transforms impedances.

11. A typical "light dimmer" used to dim the stage lights in a theater consists of a variable inductor L connected in series with the lightbulb B as shown in Fig. 37-16. The power supply is 120 V (rms) at 60.0 Hz; the lightbulb is marked "120 V, 1000 W." (*a*) What maximum inductance L is required if the power in the lightbulb is to be varied by a factor of five? Assume that the resistance of the lightbulb is independent of its temperature. (*b*) Could one use a variable resistor instead of an inductor? If so, what maximum resistance is required? Why is this not done?

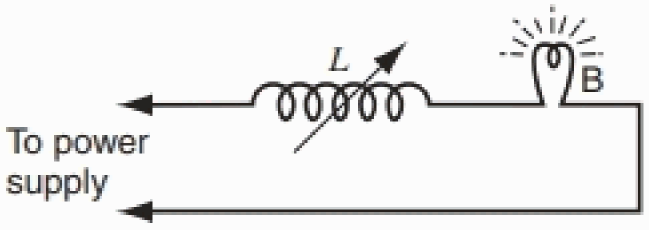


FIGURE 37-16. Problem 11.

COMPUTER PROBLEM

1. Consider Sample Problem 37-3 again. Using numerical methods based on Eq. 37-25, generate a graph of the current as a function of time. Choose a step size of 1/6000 s (1/100 of a period) and assume that at $t = 0$ the charge on the capacitor is zero and there is no current in the circuit. How much time is required for the current to settle into a solution of the form of Eq. 37-2? How do your results compare to the analytical solution?

CHAPTER 38

MAXWELL'S EQUATIONS AND ELECTROMAGNETIC WAVES

In this chapter we summarize the four basic equations of electromagnetism, known as Maxwell's equations, which we have discussed individually in previous chapters. An argument based on symmetry reveals an important term missing from one of the equations. When this missing term is included, these equations lead to the prediction of electromagnetic waves that travel with the speed of light. We discuss the properties of these waves, which are important in understanding optics, radio and TV transmission, and the properties of practical devices such as microwave ovens.

38-1 THE BASIC EQUATIONS OF ELECTROMAGNETISM

Although there are many differences in the physical properties of electric and magnetic fields, there are many similarities in their mathematical properties. To see these similarities, let us write the basic equations of electromagnetism to apply to a region of space in which electric and magnetic fields exist but there are no charges or currents (the fields may be caused by charges and currents in other regions of space).

If we choose any closed surface in this region, we can apply Gauss' law for both electric and magnetic fields:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = 0, \tag{38-1}$$

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0. \tag{38-2}$$

Over any closed surface, the surface integrals of the electric and magnetic fields are both zero, because the surface encloses no electric charge or magnetic poles. These two equations have exactly the same form, which represents an important symmetry between electric and magnetic fields.

We now choose any closed path in this region and apply Faraday's law and Ampère's law:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -\frac{d\Phi_B}{dt}, \tag{38-3}$$

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = 0. \tag{38-4}$$

The symmetry between $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ that was present in Eqs. 38-1 and 38-2 seems to be missing in Eqs. 38-3 and 38-4. Faraday's law, Eq. 38-3, tells us that in this region a varying magnetic field can set up an electric field. Is it possible that a varying *electric* field could set up a *magnetic* field?

This question was first asked by Maxwell.* His correct answer provided the missing term in Ampère's law that restored the symmetry between electric and magnetic fields in Faraday's law and Ampère's law. This additional term is responsible for the existence of electromagnetic waves, which Maxwell was able to deduce from his theory. Soon after his prediction, the waves were discovered and applied to the invention of radio. It is no exaggeration to state that the entire field of modern communication follows directly from Maxwell's discovery.

---

*James Clerk Maxwell (1831–1879), a Scottish physicist, was the first to provide the mathematical structure of the laws of electromagnetism. He also developed the statistical mechanics of gases and made important contributions to color photography and to the understanding of Saturn's rings. The equations of electromagnetism were developed in their present form not by Maxwell but by British physicist Oliver Heaviside (1850–1925), who recognized the symmetries betweeen $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ in them.

38-2 INDUCED MAGNETIC FIELDS AND THE DISPLACEMENT CURRENT

Here we discuss in detail the evidence for the supposition of the previous section—namely, that a changing electric field induces a magnetic field. Although we are guided primarily by considerations of symmetry, we also find direct experimental verification.

Figure 38-1*a* shows a circular parallel-plate capacitor. A current i enters the left-hand plate (which we assume to carry a positive charge), and an equal current i leaves the right-hand plate. An Ampèrian loop surrounds the wire in Fig. 38-1*a* and forms the boundary for a surface that is pierced by the wire. The current in the wire sets up a magnetic field; in Section 33-5 we saw that the magnetic field and the current are related by Ampère's law,

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 i. \tag{38-5}$$

That is, the line integral of the magnetic field around the loop is proportional to the total current that passes through the surface bounded by the loop.

In Fig. 38-1*b*, we have kept the same loop but have stretched the surface bounded by the loop so that it encloses the entire left-hand capacitor plate. Since the loop has not changed (nor has the magnetic field), the left side of Ampère's law gives the same result, but the right side gives a very different result—namely, zero—because no conducting wires pass through the surface. We appear to have a violation of Ampère's law!

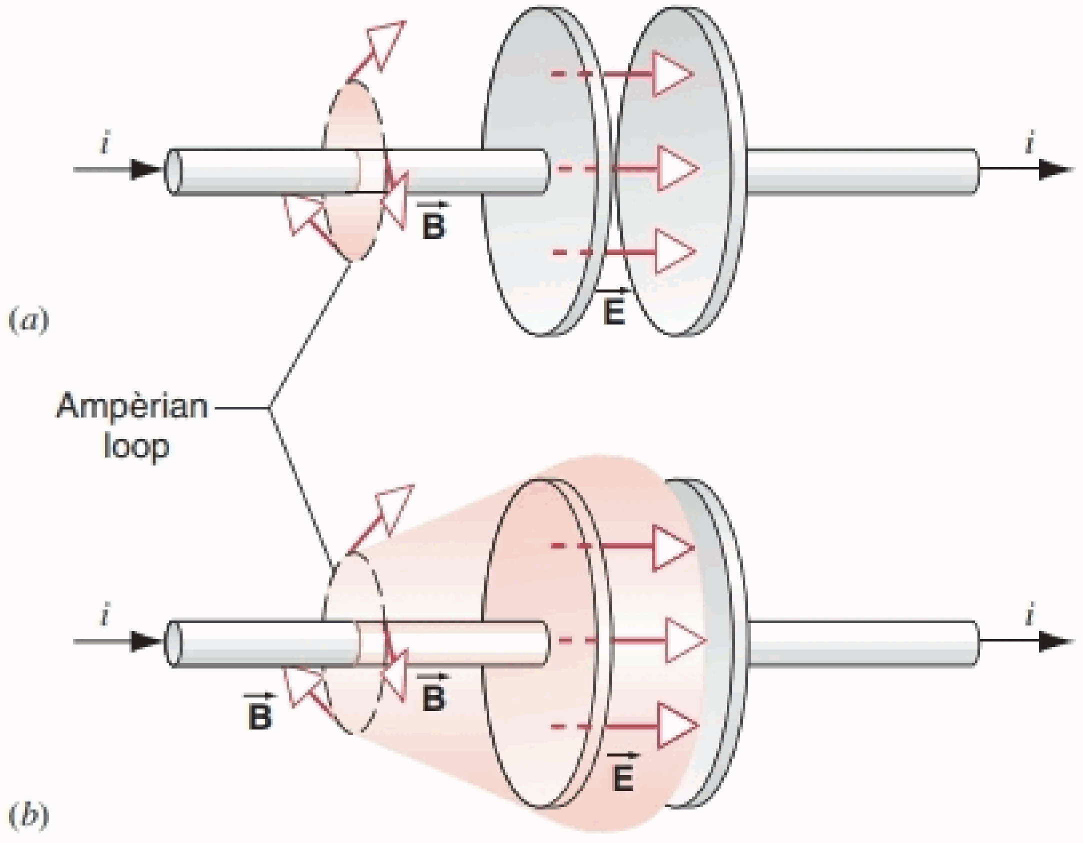


FIGURE 38-1. (*a*) An Ampèrian loop encloses a surface through which passes a wire carrying a current. (*b*) The same Ampèrian loop encloses a surface that passes between the capacitor plates. No conduction current passes through the surface.

To restore Ampère's law so that it correctly describes the situation of Fig. 38-1*b*, we rely on the conclusion given in the previous section based on symmetry: *a magnetic field is set up by a changing electric field.* Let us consider the situation of Fig. 38-1 in more detail. As charge is transported into the capacitor, the electric field in its interior changes at a certain rate dE/dt. The electric field lines pass through the surface of Fig. 38-1*b* inside the capacitor; we account for the passage of field lines through this surface in terms of the electric flux Φ_E, and a changing electric field must give a correspondingly changing electric flux, $d\Phi_E/dt$.

To describe this new effect quantitatively, we are guided by analogy with Faraday's law of induction, Eq. 38-3, which asserts that an electric field (left side) is produced by a changing magnetic field (right side). For the symmetrical counterpart we write*

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}. \tag{38-6}$$

Equation 38-6 asserts that a magnetic field (left side) can be produced by a changing electric field (right side).

The situation shown in Fig. 38-1*a* is described by Ampère's law in the form of Eq. 38-5, whereas the situation of Fig. 38-1*b* is described by Eq. 38-6. In the first case, it is the current through the surface that gives the magnetic field, and in the second case, it is the changing electric flux through the surface that gives the magnetic field. In general, we must account for *both* ways of producing a magnetic field—(*a*) by a current and (*b*) by a changing electric flux—and so we must modify Ampère's law to read

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 i + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}. \tag{38-7}$$

Maxwell is responsible for this important generalization of Ampère's law.

In Chapter 33 we assumed that no changing electric fields were present so that the term $d\Phi_E/dt$ in Eq. 38-7 was zero. In the discussion of Fig. 38-1*b* we assumed that there were no conduction currents in the space containing the electric field. Thus the term $\mu_0 i$ in Eq. 38-7 is zero in that case. We see now that each of these situations is a special case. If there were fine wires connecting the two plates in Fig. 38-1*b*, there would be contributions from both terms in Eq. 38-7.

An alternative way of interpreting Eq. 38-7 is suggested by Fig. 38-2, which shows the electric field in the region between the capacitor plates of Fig. 38-1. We now take our Ampèrian loop to be a circular path in this region. On the right side of Eq. 38-7, the term i is zero, but the term $d\Phi_E/dt$ is not zero. In fact, the flux through the surface is positive if

---

*The SI system of units requires that we insert the constants ϵ_0 and μ_0 in Eq. 38-6. In some other unit systems they would not appear.

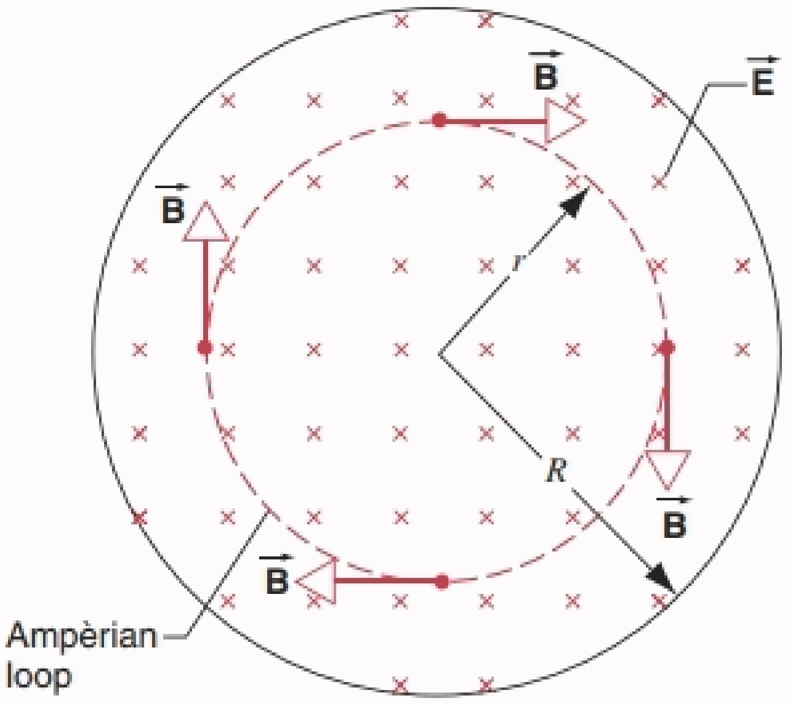


FIGURE 38-2. The induced magnetic field $\vec{\mathbf{B}}$, shown at four points, produced by the changing electric field $\vec{\mathbf{E}}$ inside the capacitor of Fig. 38-1. The electric field is increasing in magnitude. Compare with Fig. 34-15*b*.

the field lines are as shown, and the flux is increasing (corresponding to the electric field increasing) as positive charge is transported into the left-hand plate of Fig. 38-1. The line integral of $\vec{\mathbf{B}}$ around the loop must also be positive, and the directions of $\vec{\mathbf{B}}$ must be as shown in Fig. 38-2.

Figure 38-2 suggests a beautiful example of the symmetry of nature. A changing *magnetic* field induces an *electric* field (Faraday's law); now we see that a changing *electric* field induces a *magnetic* field. Carefully compare Fig. 38-2 with Fig. 34-15*b*, which illustrates the production of an electric field by a changing magnetic field. In each case the appropriate flux Φ_B or Φ_E is *increasing*. However, experiment shows that the lines of $\vec{\mathbf{E}}$ in Fig. 34-15*b* are *counterclockwise,* whereas those of $\vec{\mathbf{B}}$ in Fig. 38-2 are *clockwise.* This difference requires that the minus sign of Eq. 38-3 be omitted from Eq. 38-6.

SAMPLE PROBLEM 38-1. A parallel-plate capacitor with circular plates is being charged as in Fig. 38-1. (*a*) Derive an expression for the induced magnetic field at radius r in the region between the plates. Consider both $r \leq R$ and $r \geq R$. (*b*) Find B at $r = R$ for $dE/dt = 10^{12}$ V/m·s and $R = 5.0$ cm.

Solution (*a*) From Eq. 38-6,

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 \epsilon_0 \frac{d\Phi_E}{dt},$$

we can write, for $r \leq R$, as in Fig. 38-2,

$$(B)(2\pi r) = \mu_0 \epsilon_0 \frac{d}{dt}[(E)(\pi r^2)] = \mu_0 \epsilon_0 \pi r^2 \frac{dE}{dt}.$$

Solving for B yields

$$B = \tfrac{1}{2}\mu_0 \epsilon_0 r \frac{dE}{dt} \qquad (r \leq R).$$

For $r \geq R$, Eq. 38-6 yields

$$(B)(2\pi r) = \mu_0 \epsilon_0 \frac{d}{dt}[(E)(\pi R^2)] = \mu_0 \epsilon_0 \pi R^2 \frac{dE}{dt},$$

or

$$B = \frac{\mu_0 \epsilon_0 R^2}{2r} \frac{dE}{dt} \qquad (r \geq R).$$

(*b*) At $r = R$ the two equations for B reduce to the same expression, or

$$\begin{aligned} B &= \tfrac{1}{2}\mu_0 \epsilon_0 R \frac{dE}{dt} \\ &= \tfrac{1}{2}(4\pi \times 10^{-7}\ \mathrm{T \cdot m/A})(8.9 \times 10^{-12}\ \mathrm{C^2/N \cdot m^2}) \\ &\quad \times (5.0 \times 10^{-2}\ \mathrm{m})(10^{12}\ \mathrm{V/m \cdot s}) \\ &= 2.8 \times 10^{-7}\ \mathrm{T} = 280\ \mathrm{nT}. \end{aligned}$$

This shows that the induced magnetic fields in this example are so small that they can scarcely be measured with simple apparatus, in sharp contrast to induced *electric* fields (Faraday's law), which can be demonstrated easily. This experimental difference is in part due to the fact that induced emfs can easily be multiplied by using a coil of many turns. No technique of comparable simplicity exists for magnetic fields. In experiments involving oscillations at very high frequencies, *dE*/*dt* can be very large, resulting in significantly larger values of the induced magnetic field.

Displacement Current

Comparison of Eqs. 38-5 and 38-6 shows that the term $\epsilon_0\, d\Phi_E/dt$ has the dimensions of a current. Even though no motion of charge is involved, there are advantages in giving this term the name *displacement current.** The displacement current i_d is defined according to

$$i_\mathrm{d} = \epsilon_0 \frac{d\Phi_E}{dt}. \tag{38-8}$$

Thus we can say that a magnetic field can be set up either by a conduction current i or by a displacement current i_d, and we can rewrite Eq. 38-6 as

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 (i + i_\mathrm{d}). \tag{38-9}$$

Let us calculate the displacement current i_d in the capacitor gap of Fig. 38-1*b*. The charge q on the plates is related to the electric field E in the gap by Eq. 30-3 ($E = \sigma/\epsilon_0$),

$$q = \epsilon_0 EA.$$

Differentiating gives

$$i = \frac{dq}{dt} = \epsilon_0 \frac{d(EA)}{dt}.$$

The quantity EA is the electric flux Φ_E, and thus

$$i = \epsilon_0 \frac{d\Phi_E}{dt}.$$

---

*The word "displacement" was introduced for historical reasons. It has nothing to do with our previous use of displacement to indicate the position of a particle.

Comparison with Eq. 38-8 shows

$$i = i_d.$$

Thus the displacement current in the gap equals the conduction current in the wires.

The concept of displacement current thus permits us to retain the notion that *current is continuous,* a principle established for steady conduction currents in Section 31-1. In Fig. 38-1*b*, for example, a conduction current i enters the positive plate and leaves the negative plate. The *conduction* current is *not* continuous across the capacitor gap because no charge is transported across this gap. However, the displacement current i_d in the gap is exactly equal to i, thus retaining the concept of the continuity of current.

When the capacitor is fully charged, the conduction current drops to zero (no current flows in the wires). The electric field between the plates becomes constant; thus $dE/dt = 0$, and so the displacement current also drops to zero.

The displacement current i_d, given by Eq. 38-8, has a direction as well as a magnitude. The direction of the conduction current i is that of the conduction current density vector $\vec{\mathbf{j}}$. Similarly, the direction of the displacement current i_d is that of the displacement current density vector $\vec{\mathbf{j}}_\text{d}$, which, as we deduce from Eq. 38-8, is $\epsilon_0(d\vec{\mathbf{E}}/dt)$. The right-hand rule applied to i_d gives the direction of the associated magnetic field, just as it does for the conduction current i.

SAMPLE PROBLEM 38-2. What is the displacement current for the situation of Sample Problem 38-1?

Solution From Eq. 38-8, the definition of displacement current,

$$i_\text{d} = \epsilon_0 \frac{d\Phi_E}{dt} = \epsilon_0 \frac{d}{dt}[(E)(\pi R^2)] = \epsilon_0 \pi R^2 \frac{dE}{dt}$$

$$= (8.9 \times 10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2)(\pi)(5 \times 10^{-2}\ \text{m})^2(10^{12}\ \text{V/m}\cdot\text{s})$$

$$= 0.070\ \text{A} = 70\ \text{mA}.$$

This is a reasonably large current, yet we determined in Sample Problem 38-1 that it produces a magnetic field of only 280 nT. A current of 70 mA flowing in a thin wire would produce a large magnetic field near the surface of the wire, easily detectable by a compass needle.

The difference is *not* caused by the fact that one current is a conduction current and the other is a displacement current. Under the same conditions, both kinds of current are equally effective in generating a magnetic field. The difference arises because the conduction current, in this case, is confined to a thin wire but the displacement current is spread out over an area equal to the surface area of the capacitor plates. Thus the capacitor behaves like a "fat wire" of radius 5 cm, carrying a (displacement) current of 70 mA. Its largest magnetic effect, which occurs at the capacitor edge, is much smaller than would be the case at the surface of a thin wire. (See also Problem 2.)

38-3 MAXWELL'S EQUATIONS

With the addition of the induced magnetic field (displacement current) term in Ampère's law, the four basic equations of electromagnetism, known as Maxwell's equations, are now complete. These equations are summarized in Table 38-1, which also shows the crucial experiment that led to each equation. This list of experiments reminds us that Maxwell's equations are not mere theoretical speculations; rather, each was developed to explain the results of laboratory experiments.

There are several features of these remarkable equations that we should consider.

1. *Symmetry.* The inclusion of the displacement current certainly makes Eqs. III and IV of Table 38-1 look more similar, thereby improving the symmetry of the equations. If individual magnetic charges (magnetic monopoles) were confirmed to exist, the equations would look even more similar. If q_m represents the "magnetic charge," then we would write Gauss' law for magnetism as $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = \mu_0 q_\text{m}$. This equation would assert that the magnetic flux through any closed surface is proportional to the net magnetic charge enclosed by the surface. In this case Eqs. I and II would become more symmetric.

If we could form a stream of these magnetic charges into a magnetic current $i_\text{m} = dq_\text{m}/dt$, then (assuming that our experiments verified this) we could add a term to the right side of Eq. III showing that a magnetic current can set up an electric field. This would make Eqs. III and IV more symmetric.

So far there is no experimental evidence for the existence of magnetic monopoles. However, if they were to be discovered it would take little effort to modify Maxwell's equations to account for their effects.

2. *Electromagnetic waves.* The four individual equations were known before Maxwell's time, and apart from the displacement current no new predictions arise from any individual equation. As we show in the next section, when the equations are combined, a new prediction does emerge—the existence of electromagnetic waves and a value for their speed (the speed of light). These waves were predicted by Maxwell and discovered by Heinrich Hertz in 1888, 15 years after Maxwell's theory was published. Later in this chapter we show how this prediction results from Maxwell's equations.

3. *Electromagnetism and relativity.* What is especially remarkable about Maxwell's equations is that they are entirely consistent with the special theory of relativity. In contrast to Newton's laws for mechanics, which require major changes for motion at speeds close to the speed of light, Maxwell's equations remain the same for all observers, no matter what their relative speeds. In fact, Einstein's discovery of relativity grew directly from his thinking about the laws of electromagnetism and Maxwell's equations.

TABLE 38-1 **Basic Equations of Electromagnetism (Maxwell's Equations)**[a]

Number	*Name*	*Equation*	*Describes*	*Crucial Experiment*	*Chapter Reference*
I	Gauss' law for electricity	$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q/\epsilon_0$	Charge and the electric field	(*a*) Like charges repel and unlike charges attract, as the inverse square of their separation. (*b*) A charge on an insulated conductor moves to its outer surface.	27
II	Gauss' law for magnetism	$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0$	The magnetic field	Magnetic field lines form closed loops; there is no evidence that magnetic monopoles exist.	35
III	Faraday's law of induction	$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -d\Phi_B/dt$	The electrical effect of a changing magnetic field	A bar magnet, thrust through a closed loop of wire, will set up a current in the loop.	34
IV	Ampère's law (as extended by Maxwell)	$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 i + \mu_0\epsilon_0\, d\Phi_E/dt$	The magnetic effect of a current or a changing electric field	(*a*) A current in a wire sets up a magnetic field near the wire. (*b*) The speed of light can be calculated from purely electromagnetic measurements.	33 38

[a] Written on the assumption that no dielectric or magnetic material is present.

Cavity Oscillations (Optional)

There are many situations in which Maxwell's equations provide a theoretical understanding of a practical device or phenomenon. One particular case is the metal cavity containing oscillating electric and magnetic fields.

The electromagnetic cavity oscillator is similar in many ways to an acoustic cavity oscillator, such as a closed organ pipe. When the pipe is set into oscillation by an external disturbance, we can produce a standing wave with oscillations in the density and velocity of the air molecules, such that the acoustic energy in the pipe oscillates between the potential energy associated with compressions and rarefactions of the gas and the kinetic energy of the moving gas.

The electromagnetic resonant cavity behaves in a similar way, except that the energy oscillates between its magnetic and electric fields in an electromagnetic standing wave. Figure 38-3 shows a representation of the electric and magnetic fields in one half-cycle of oscillation. Electric and magnetic fields occur throughout the volume of the cavity. The energy densities at any point are given by Eqs. 30-28 ($u_E = \frac{1}{2}\epsilon_0 E^2$) and 36-22 ($u_B = B^2/2\mu_0$), and the total stored energies are found by integrating over the volume of the cavity.

We can think of the two flat ends of the cavity as the plates of a capacitor that instantaneously hold charges $+q$ and $-q$ to establish the electric field shown in Fig. 38-3. As the cavity oscillates, the charges travel along the conducting walls of the cavity to reverse the field direction.

If we choose a circular path in the cavity in a plane parallel to the ends, then the electric flux through the area surrounded by the path is changing. This changing electric flux sets up a magnetic field tangent to the circle, according to the extra term Maxwell added to Ampère's law. Equivalently, we can regard displacement current as flowing inside

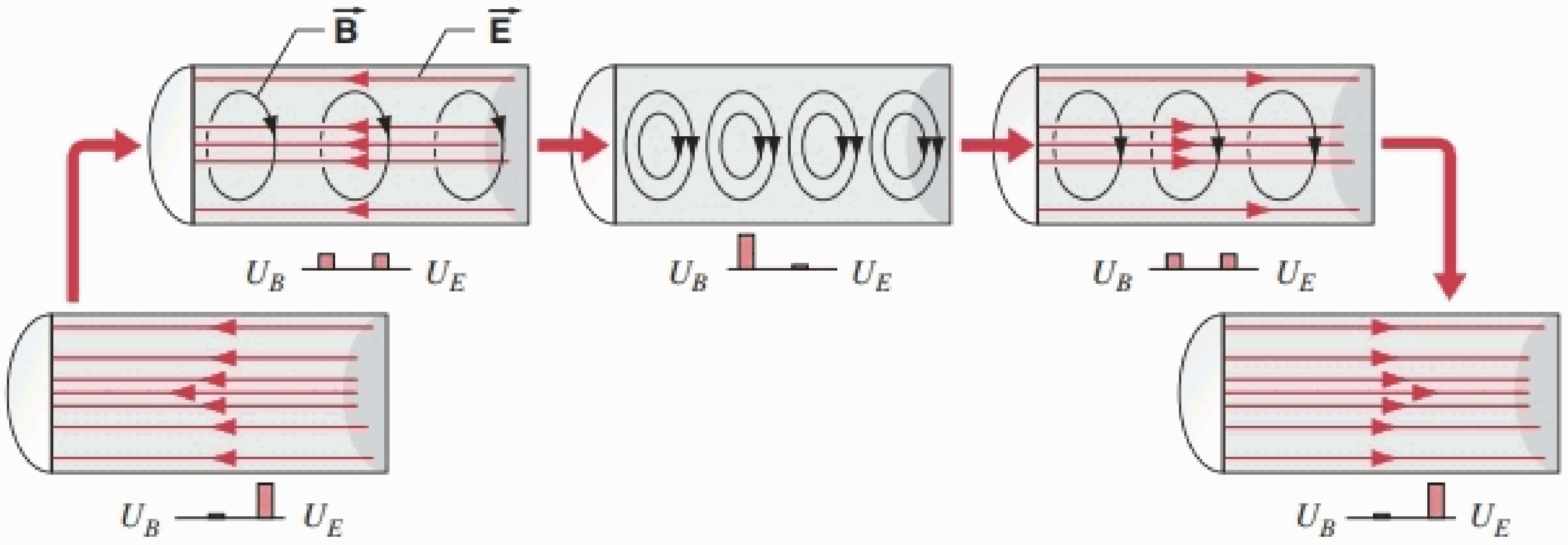


FIGURE 38-3. One half-cycle of oscillation of an electromagnetic cavity. The graphs below each figure show the stored electric energy U_E and magnetic energy U_B. The lines of $\vec{\mathbf{E}}$ are parallel to the axis of the cavity, and the lines of $\vec{\mathbf{B}}$ are circles concentric with the axis.

the volume of the cavity (parallel to the electric field lines), and this displacement current sets up circular magnetic field lines just like ordinary currents in straight wires.

If we instead choose a rectangular path in the plane of the electric field lines, Faraday's law shows that the electric field on that path depends on the rate of change of magnetic flux through the rectangle. So we have two results from Maxwell's equations:

$$B \propto \frac{d\Phi_E}{dt} \quad \text{and} \quad E \propto \frac{d\Phi_B}{dt}. \qquad (38\text{-}10)$$

Note that B has its maximum value when the electric field is changing the most rapidly, which occurs when $E = 0$, that is, when the electric field is reversing direction. Similarly, E has its maximum value when $B = 0$.

Equations 38-10 show the interdependence of $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ in the cavity. The changing electric field gives a magnetic field, and the changing magnetic field gives an electric field. The oscillations of $\vec{\mathbf{E}}$ produce $\vec{\mathbf{B}}$ and the oscillations of $\vec{\mathbf{B}}$ produce $\vec{\mathbf{E}}$. The oscillating fields thus sustain each other, and once the oscillation is established it would continue indefinitely were it not for resistive losses in the cavity walls or leakage of energy through openings in the cavity. A similar mutual coupling of electric and magnetic fields occurs in electromagnetic traveling waves, which we discuss later in this chapter.

Oscillating cavities such as we have described form the basis for the *magnetron,* which served as a generator of microwave radiation for use in radar during the 1940s. Another cavity-based device was the *klystron,* which was used to amplify reflected radar signals. (If the waves travel through a klystron cavity at the speed of light, you can estimate that a cavity a few centimeters in length will have a period of about 10^{-10} s, corresponding to a frequency of 10 GHz.) A common use of the klystron is in accelerators that produce beams of charged particles with high energy. Figure 38-4 shows the interior of the two-mile electron accelerator at Stanford, in which hundreds of resonant cavities in succession are used to boost the energies of the electrons to 50 GeV. ■

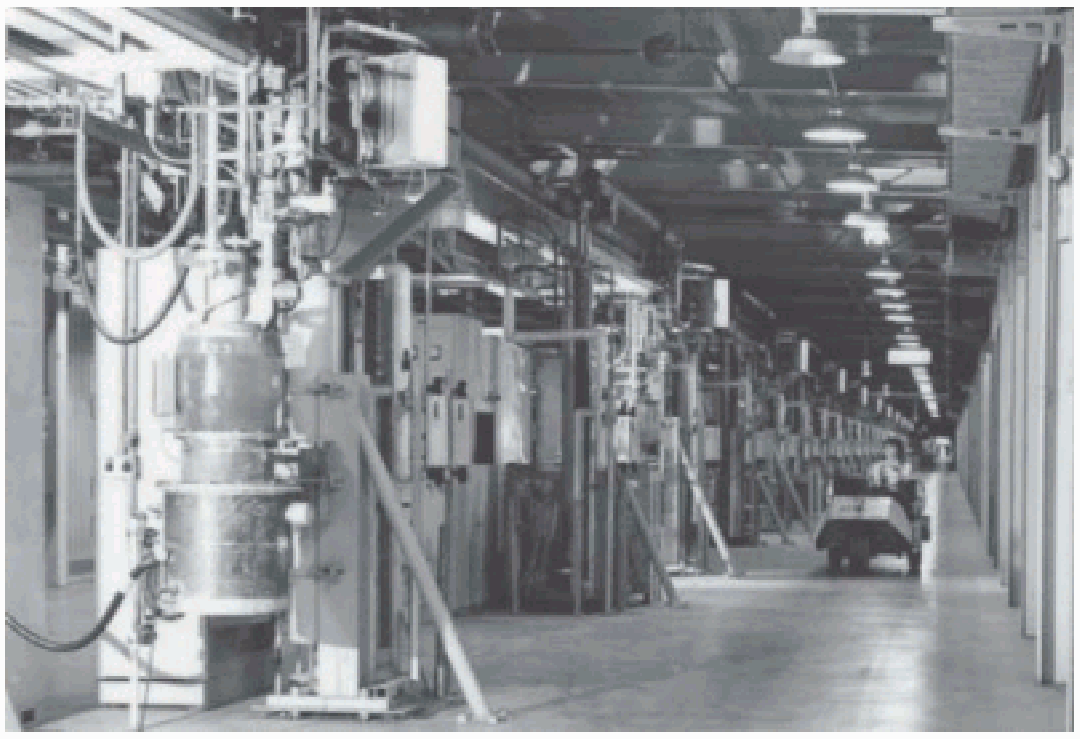


FIGURE 38-4. The interior of the 2-mile Stanford Linear Accelerator. The large vertical cylinder is one of the several hundred electromagnetic resonant cavities (klystrons) that supply the electric fields needed to accelerate the electrons. Each klystron produces a peak power of 67 MW.

38-4 GENERATING AN ELECTROMAGNETIC WAVE

Electromagnetic waves carry energy or momentum from one point in space to another by means of their electric and magnetic fields. Before we consider the mathematical description of electromagnetic waves, let us discuss some general properties that we might expect for these waves.

What kind of arrangement of charges or currents would we expect to produce an electromagnetic wave? An electric charge at rest sets up a pattern of electric field lines. A charge in motion at constant speed sets up a pattern of magnetic field lines in addition to the electric field lines. Once a steady condition has been reached (that is, after the charge is in motion and the fields are established in space), there is an energy density in space associated with the electric and magnetic fields, but the energy density remains constant in time. No signal, other than evidence of its presence, is transported from the charge to distant points; there is no transport of energy or momentum, and there is no electromagnetic radiation.

If, on the other hand, you were to wiggle the charge back and forth, you could send signals to a distant friend who had the equipment necessary to detect changes in the electric and magnetic fields. With a prearranged code, you could send information by wiggling the charge at a certain rate or in a certain direction. In this case, you would be signaling by means of an electromagnetic wave. To produce this wave, it is necessary to accelerate the charge. That is, *static charges and charges in motion at constant velocity do not radiate; accelerated charges radiate.* Put another way, the uniform motion of charges is equivalent to a current that does not change with time, and the accelerated motion of charges is correspondingly a current that varies with time; thus we can equivalently regard radiation as being produced by time-varying currents.

In the laboratory, a convenient way of generating an electromagnetic wave is to cause currents in wires to vary with time. We assume for simplicity a sinusoidal time variation. Figure 38-5 shows a circuit that might be used for this purpose. It consists of an oscillating *RLC* circuit, with an external source that restores the energy that is dissipated in the circuit or carried away by the radiation. The current in the circuit varies sinusoidally with the resonant angular frequency ω, which is approximately $1/\sqrt{LC}$ if the resistive losses are small (see Section 36-7). The oscillator is coupled through a transformer to a *transmission line,* which serves to carry the current to an *antenna.* (Coaxial cables, which carry TV signals to many homes, are common examples of transmission lines.)

The geometry of the antenna determines the geometrical properties of the radiated electric and magnetic fields. We assume a *dipole antenna,* which, as Fig. 38-5 shows, can be considered simply as two straight conductors. Charges

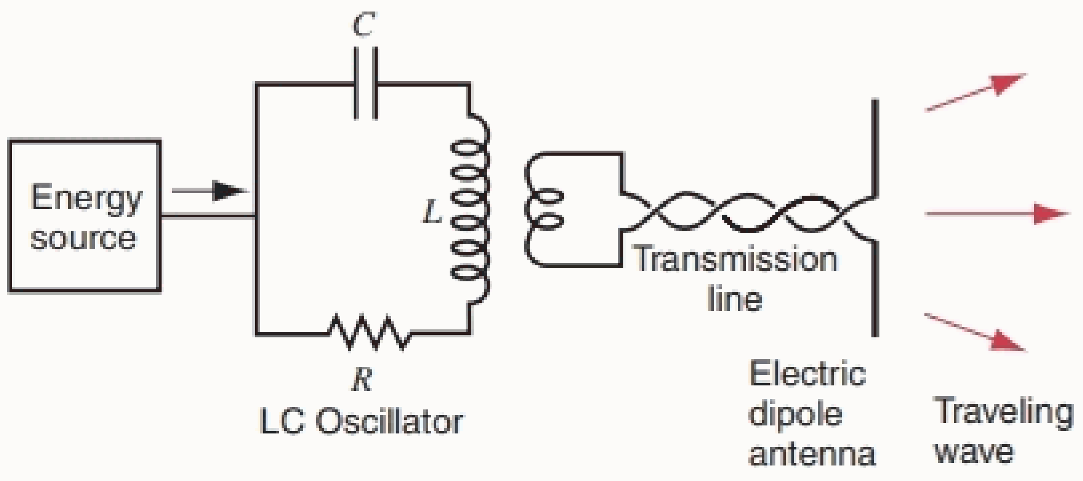


FIGURE 38-5. An arrangement for generating a traveling electromagnetic wave.

surge back and forth in these two conductors at the frequency ω, driven by the oscillator. We can regard the antenna as an oscillating electric dipole, in which one branch carries an instantaneous charge q and the other branch carries $-q$. The charge q varies sinusoidally with time and changes sign every half cycle. The charges are certainly accelerated as they move back and forth in the antenna, and as a result the antenna is a source of *electric dipole radiation.* At any point in space there are electric and magnetic fields that vary sinusoidally with time.*

Figure 38-6 shows a series of "snapshots" that give a schematic picture of how the radiation field is formed. The electric field lines can be deduced from the locations of the positive and negative charges of the dipole; the corresponding magnetic field, shown in Fig. 38-6*e*, can be inferred from the current in the conductors using the right-hand rule. The figure is a slice through the xy plane; to obtain a more complete picture of the field, we must imagine the figure to be rotated about the y axis. We assume that we observe the field at distances from the dipole that are large compared with its dimensions and compared with the wavelength of the radiation; the field observed under these conditions is called the *radiation field.* At smaller distances, we would observe the more complicated *near field,* which we do not discuss here. Note that the field "breaks away" from the antenna and forms closed loops, in contrast to the static field of an electric dipole, in which the field lines always start on positive charges and end on negative charges.

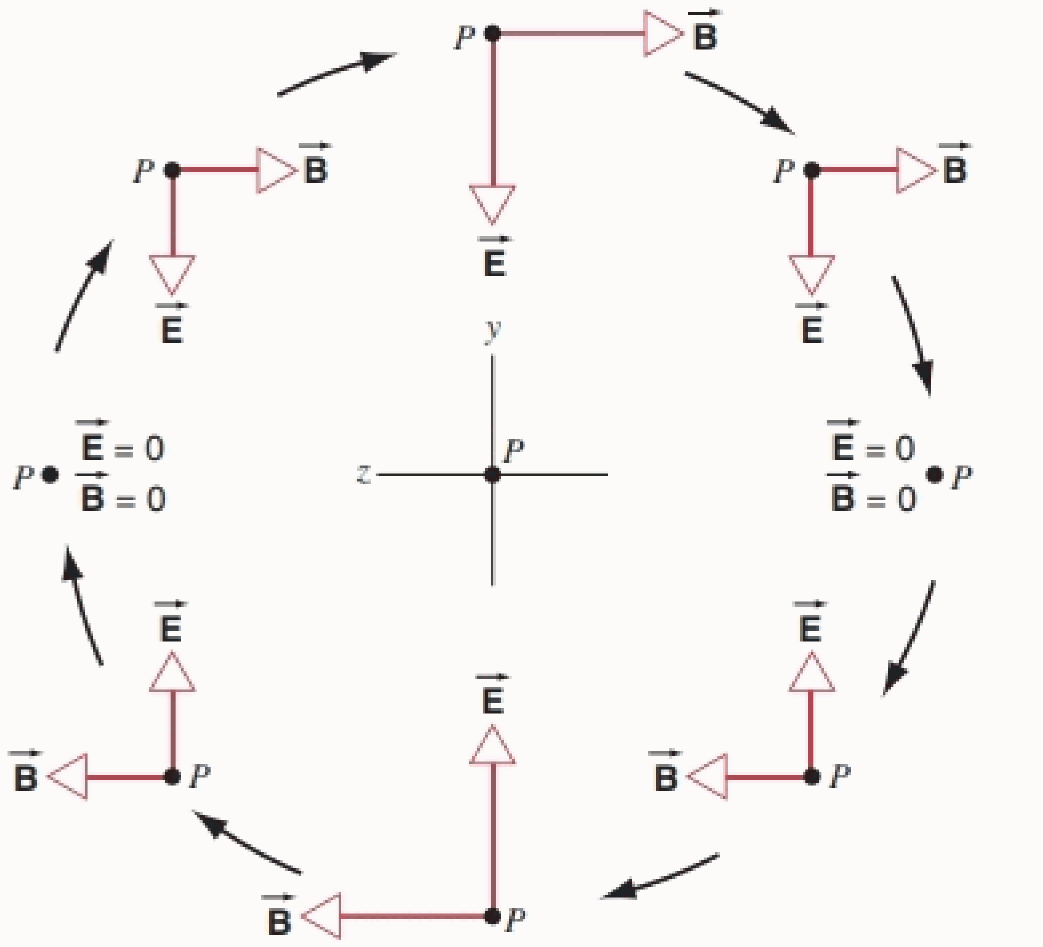


FIGURE 38-7. Eight cyclical "snapshots" of the plane electromagnetic wave radiated from the oscillating dipole of Fig. 38-6 observed at point P. The direction of travel of the wave (x direction in Fig. 38-6) is out of the plane of the page. Lines of $\vec{E}$ are vertical, and lines of $\vec{B}$ are horizontal.

An alternative view of the radiation field is given in Fig. 38-7, which represents a series of "snapshots" of the electric and magnetic fields sweeping past an observer located at point P on the x axis of Fig. 38-6. We assume the observer to be located so far from the dipole that the wavefronts can be regarded as planes. As is always the case, the density of field lines indicates the strength of the field. Note especially that (1) $\vec{E}$ and $\vec{B}$ are in phase (they both reach their maxima at the same instant, and they both are zero at the same instant), and (2) $\vec{E}$ and $\vec{B}$ are perpendicular to one

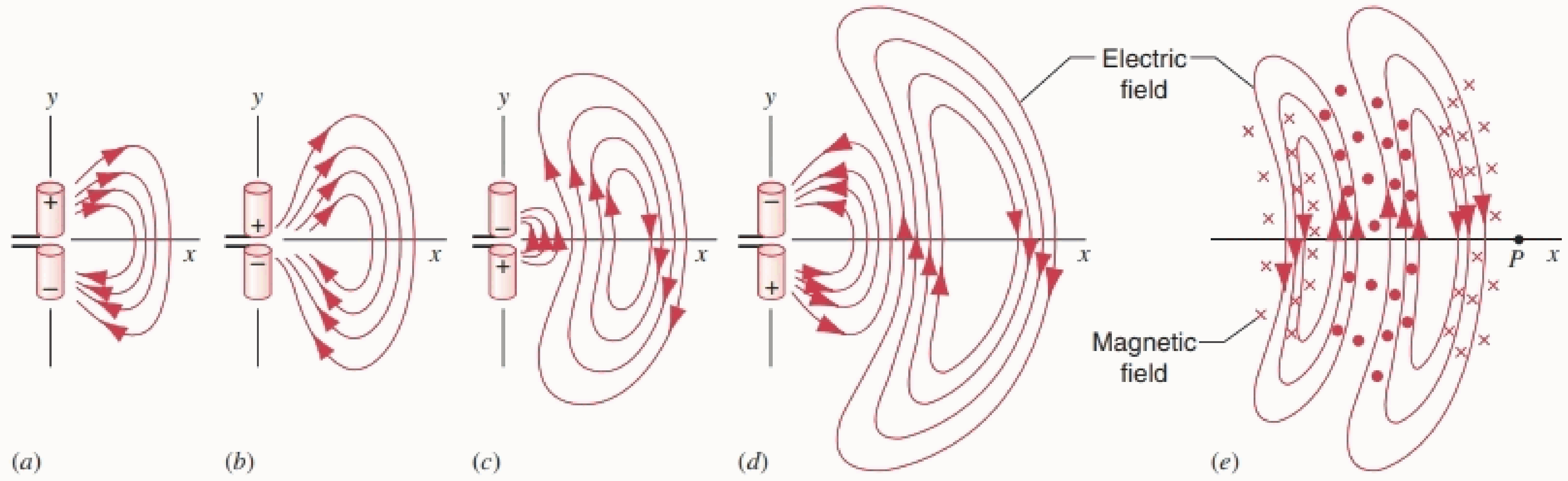


FIGURE 38-6. Successive stages in the emission of an electromagnetic wave from a dipole antenna. In (*a*)–(*d*), only the electric field patterns are shown. In (*e*), the magnetic field is shown as perpendicular to the plane of the page.

---

* Most of the radiations we encounter, from radio waves to light to x rays and gamma rays, are of the dipole type. Radio and TV antennas are generally designed to transmit dipole radiation. Individual atoms and nuclei can often be considered as oscillating dipoles from the standpoint of emitting radiation.

another. These conclusions, which follow from an analysis of traveling electromagnetic waves in free space using Maxwell's equations, are discussed in the next section.

38-5 TRAVELING WAVES AND MAXWELL'S EQUATIONS

The preceding discussion has given us a qualitative picture of one type of electromagnetic traveling wave. In this section we consider the mathematical description of the wave, which we show to be consistent with Maxwell's equations. In doing so, we also show that the speed of such waves in empty space is the same as the speed of light, which leads us to conclude that light is itself an electromagnetic wave.

Suppose the observer at point *P* in Fig. 38-6 is at such a great distance from the oscillating dipole that the wavefronts passing point *P* (shown in Fig. 38-7) are planes. The lines of $\vec{\mathbf{E}}$ are parallel to the *y* axis, and the lines of $\vec{\mathbf{B}}$ are parallel to the *z* axis. We write the $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ fields in the usual mathematical form of a sinusoidal traveling wave (see Section 18-3):

$$E(x, t) = E_{\mathrm{m}} \sin (kx - \omega t), \tag{38-11}$$

$$B(x, t) = B_{\mathrm{m}} \sin (kx - \omega t). \tag{38-12}$$

Here ω is the angular frequency associated with the oscillating dipole, and the wave number k has its usual meaning of $2\pi/\lambda$. If the wave propagates with phase speed c, then ω and k are related according to $c = \omega/k$. Figure 38-8 represents the sinusoidal variation of the $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ fields for points along the *x* axis at a particular instant of time. In this plane wave, the same variation of $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ fields occurs along any line parallel to the *x* axis; that is, the fields at any point on the *x* axis are the same as the fields everywhere in a plane through that point perpendicular to the *x* axis.

The amplitudes E_{m} and B_{m} will later be shown to be related to one another. Note that in writing these equations for the magnitudes of $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ we have assumed that E and B are in phase; that is, the phase constants in Eqs. 38-11 and 38-12 have the same value (which we have taken to be zero). Later we show that this choice follows from Maxwell's equations.

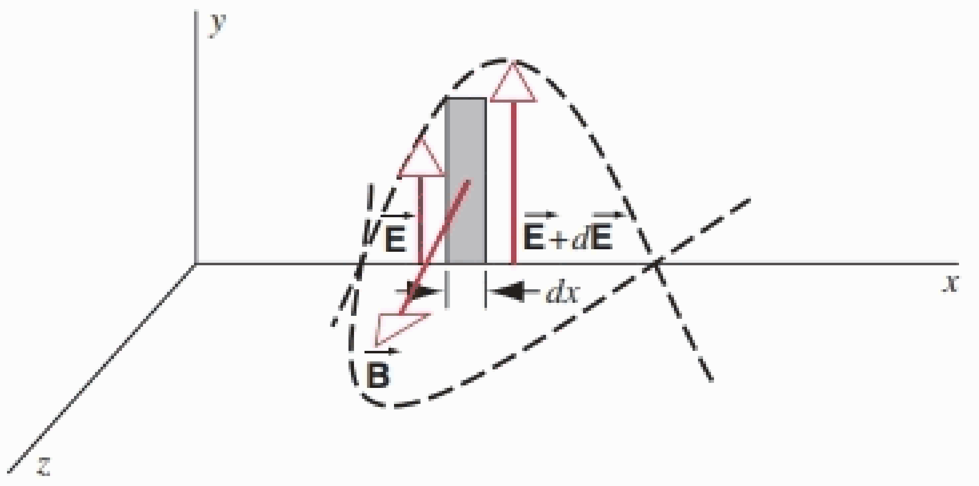


FIGURE 38-9. Detail of vertical strip from Fig. 38-8 as the wave passes through it.

We will look in detail at the wave as it passes through the two rectangular strips at point *P* in Fig. 38-8. The strips each have height h and width dx; one strip lies in the *xy* plane (that is, in the plane of $\vec{\mathbf{E}}$), and the other lies in the *xz* plane (the plane of $\vec{\mathbf{B}}$).

Let us first consider the strip in the *xy* plane, which is shown in detail in Fig. 38-9. As the wave passes over the strip, the lines of $\vec{\mathbf{E}}$ are parallel to the long sides of the strip and the lines of $\vec{\mathbf{B}}$ are perpendicular to the area of the strip. In the language of Faraday's law, as the wave passes, the magnetic flux through the rectangular area changes, and as a result there is an induced electric field around the area. This induced electric field is simply the electric field of the traveling wave.

We consider this strip at the instant of time shown in Fig. 38-9. As the wave moves to the right, the magnetic flux is decreasing with time because the $\vec{\mathbf{B}}$ field of the wave moving into the strip is smaller. The induced $\vec{\mathbf{E}}$ field must oppose this change, which means that if we imagine the boundary of the shaded rectangle to be a current loop, a counterclockwise current would be induced. This current would induce a magnetic field that would point out of the page within the rectangle. There is of course no current loop, but the $\vec{\mathbf{E}}$ vectors in Fig. 38-9 are consistent with this explanation, because the larger electric field on the right edge of the loop would give a net counterclockwise current.

Applying Faraday's law, $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -d\Phi_B/dt$, we first need the line integral of $\vec{\mathbf{E}}$ around the loop. We will evaluate the integral by traveling counterclockwise around the

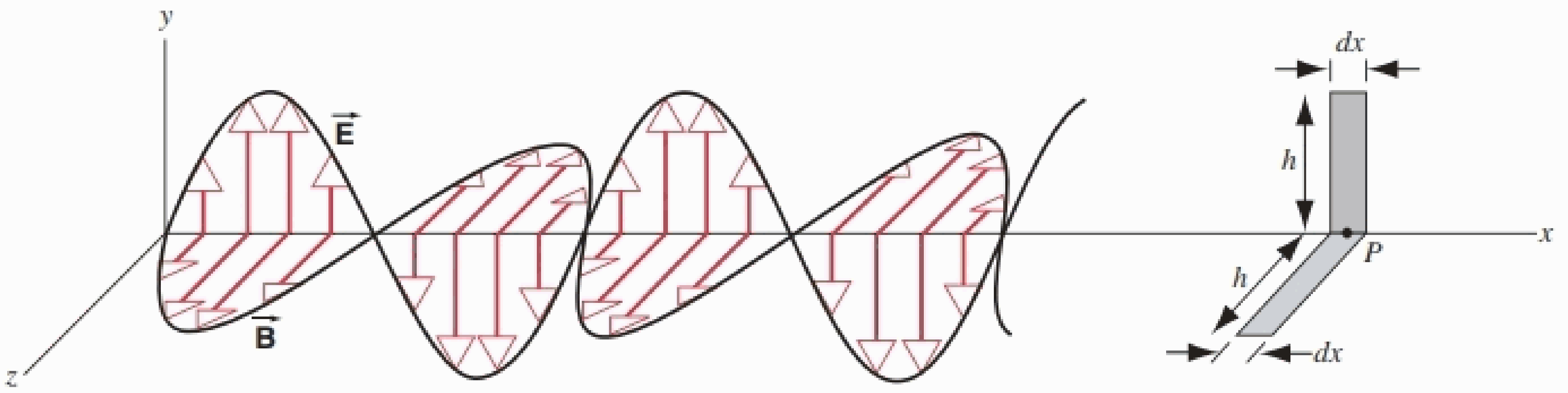


FIGURE 38-8. A "snapshot" of a sinusoidally varying electromagnetic wave traveling in the *x* direction. We consider the wave as it passes point *P*. The lengths of the $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ vectors indicate only their spatial variation at different locations on the *x* axis. The absolute lengths of the vectors are arbitrary.

loop. Because $\vec{\mathbf{E}}$ and $d\vec{\mathbf{s}}$ are perpendicular on the top and bottom of the loop, there are no contributions to the integral from those sides. The integral then becomes

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = (E + dE)h - Eh = dE\, h.$$

The flux Φ_B for the rectangle is*

$$\Phi_B = (B)(dx\, h),$$

where B is the magnitude of $\vec{\mathbf{B}}$ at the rectangular strip and $dx\, h$ is the area of the strip. Differentiating gives

$$\frac{d\Phi_B}{dt} = h\, dx \frac{dB}{dt}.$$

From Faraday's law, Eq. 38-3, we then have

$$dE\, h = -h\, dx \frac{dB}{dt},$$

or

$$\frac{dE}{dx} = -\frac{dB}{dt}. \qquad (38\text{-}13)$$

Actually, both B and E are functions of x and t; see Eqs. 38-11 and 38-12. In evaluating dE/dx, we assume that t is constant because Fig. 38-9 is an "instantaneous snapshot." Also, in evaluating dB/dt we assume that x is constant since what is required is the time rate of change of B at a particular place, the strip in Fig. 38-9. The derivatives under these circumstances are *partial derivatives,*† and a somewhat different notation is used for them; see, for example, Sections 18-3 and 18-5. In this notation, Eq. 38-13 becomes

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t}. \qquad (38\text{-}14)$$

The negative sign in this equation is appropriate and necessary, for, although E is increasing with x at the site of the shaded rectangle in Fig. 38-9, B is decreasing with t. Since $E(x, t)$ and $B(x, t)$ are known (see Eqs. 38-11 and 38-12), Eq. 38-14 reduces to

$$kE_\mathrm{m} \cos(kx - \omega t) = \omega B_\mathrm{m} \cos(kx - \omega t).$$

If we had used different phase constants in Eqs. 38-11 and 38-12, the cosine terms in this equation would be out of phase, and the two sides could not be equal at all x and t. Equation 38-14, which follows directly from applying Maxwell's equations, shows that E and B must be in phase.

Eliminating the cosine term, we obtain

$$\frac{E_\mathrm{m}}{B_\mathrm{m}} = \frac{\omega}{k} = c. \qquad (38\text{-}15)$$

The ratio of the amplitudes of the electric and the magnetic components of the wave is the speed c of the wave. From Eqs. 38-11 and 38-12 we see that the ratio of the amplitudes is the same as the ratio of the instantaneous values, or

$$E = cB. \qquad (38\text{-}16)$$

This important result will be useful in later sections.

We now turn our attention to the horizontal strip in Fig. 38-8, which lies in the xz plane. A detail of this strip is shown in Fig. 38-10. At this particular instant of time, the magnetic field $\vec{\mathbf{B}}$ increases across the strip. The electric flux is decreasing with time, because the wave moving into the strip has a smaller $\vec{\mathbf{E}}$ field.

To analyze the fields in this strip, we need the form of Ampère's law with Maxwell's modification, Eq. 38-6: $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0\epsilon_0\, d\Phi_E/dt$. In effect, the changing electric flux (or displacement current) through the rectangular area induces a magnetic field in the strip; this magnetic field is the magnetic field of the wave.

As we did in the case of the $\vec{\mathbf{E}}$ field in Fig. 38-9, let us evaluate the line integral of $\vec{\mathbf{B}}$ (the left-hand side of the Ampère–Maxwell law) by traveling counterclockwise around the strip. Because $\vec{\mathbf{B}}$ and $d\vec{\mathbf{s}}$ are perpendicular along the short edges of the strip, we need consider only its long sides, where we have

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = -(B + dB)h + Bh = -h\, dB,$$

where B is the magnitude of $\vec{\mathbf{B}}$ at the left edge of the strip and $B + dB$ is its magnitude at the right edge.

Note that $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}$ in Fig. 38-9 is positive, whereas $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}}$ in Fig. 38-10 is negative, even though the corresponding fluxes are *increasing* in both situations. This is consistent with the difference in sign in Eqs. 38-3 and 38-6 and with the figures, which show $\vec{\mathbf{E}}$ in Fig. 38-9 and $\vec{\mathbf{B}}$ in Fig. 38-10 circulating in opposite directions around the strips.

The flux Φ_E through the rectangle of Fig. 38-10 is

$$\Phi_E = (E)(h\, dx).$$

Differentiating gives

$$\frac{d\Phi_E}{dt} = h\, dx \frac{dE}{dt}.$$

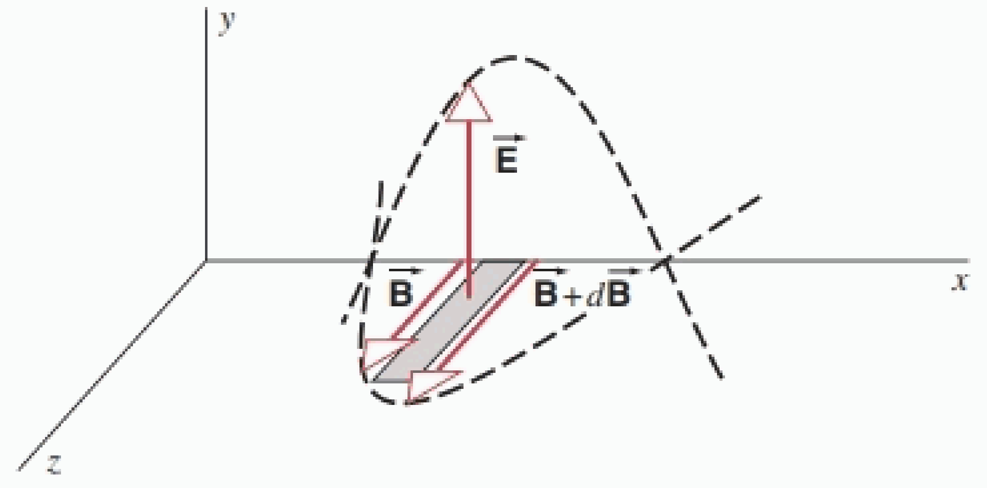


FIGURE 38-10. Detail of horizontal strip from Fig. 38-8.

---

[*] We use a right-hand rule for the sign of the flux: if the fingers of the right hand point in the direction in which we integrate around the path, then the thumb indicates the direction in which the field through the enclosed area gives a positive flux.

[†] In taking a partial derivative with respect to a certain variable, such as $\partial E/\partial x$, we treat all other variables (for instance, y, z, and t) as if they were constants.

Thus we can write the Ampère–Maxwell law, Eq. 38-6, as

$$-h\,dB = \mu_0\epsilon_0\left(h\,dx\,\frac{dE}{dt}\right)$$

or, canceling h and substituting partial derivatives,

$$-\frac{\partial B}{\partial x} = \mu_0\epsilon_0\,\frac{\partial E}{\partial t}. \qquad (38\text{-}17)$$

Again, the negative sign in this equation is appropriate and necessary, for, although B is increasing with x at the site of the shaded rectangle in Fig. 38-10, E is decreasing with t.

Combining this equation with Eqs. 38-11 and 38-12, we find

$$-kB_m\cos(kx-\omega t) = -\mu_0\epsilon_0\omega E_m\cos(kx-\omega t),$$

or

$$\frac{E_m}{B_m} = \frac{k}{\mu_0\epsilon_0\omega} = \frac{1}{\mu_0\epsilon_0 c}, \qquad (38\text{-}18)$$

where we have used $\omega = ck$. Eliminating E_m/B_m between Eqs. 38-15 and 38-18 gives

$$c = \frac{1}{\sqrt{\mu_0\epsilon_0}}. \qquad (38\text{-}19)$$

Substituting numerical values, we obtain

$$c = \frac{1}{\sqrt{(4\pi\times10^{-7}\ \mathrm{T\cdot m/A})(8.9\times10^{-12}\ \mathrm{C^2/N\cdot m^2})}}$$
$$= 3.0\times10^{8}\ \mathrm{m/s},$$

which is the speed of light in free space! This emergence of the speed of light from purely electromagnetic considerations is a crowning achievement of Maxwell's electromagnetic theory. Maxwell made this prediction before radio waves were known and before it was realized that light was electromagnetic in nature. His prediction led to the concept of the electromagnetic spectrum and to the discovery of radio waves by Heinrich Hertz in 1890. It permitted optics to be discussed as a branch of electromagnetism and allowed its fundamental laws to be derived from Maxwell's equations.

Because μ_0 is defined to be exactly $4\pi\times10^{-7}$ H/m, and the speed of light is now given the exact value of 299,792,458 m/s, Eq. 38-19 permits us to obtain a defined value of ϵ_0:

$$\epsilon_0 = \frac{1}{c^2\mu_0} = 8.854187817\ldots\times10^{-12}\ \mathrm{C^2/N\cdot m^2}.$$

Curiously, Maxwell himself did not view the propagation of electromagnetic waves and electromagnetic phenomena in general, in anything like the terms suggested by, say, Fig. 38-8. Like all physicists of his day he believed firmly that space was permeated by a subtle substance called the *ether* and that electromagnetic phenomena could be accounted for in terms of rotating vortices in this ether.

It is a tribute to Maxwell's genius that, even with such mechanical models in his mind, he was able to deduce the laws of electromagnetism that bear his name. These laws, as we have pointed out, not only required no change when Einstein's special theory of relativity came on the scene three decades later but, indeed, were strongly supported by that theory. Today, as discussed in Section 41-6, we no longer find it necessary to invoke the ether concept to explain the propagation of electromagnetic waves.

38-6 ENERGY TRANSPORT AND THE POYNTING VECTOR

Like any form of wave, an electromagnetic wave can transport energy from one location to another. Light from a bulb and radiant heat from a fire are common examples of energy flowing by means of electromagnetic waves.

The energy flow in an electromagnetic wave is commonly measured in terms of the rate of energy flow per unit area (or, equivalently, electromagnetic power per unit area). We describe the magnitude and direction of the energy flow in terms of a vector called the *Poynting vector** $\vec{\mathbf{S}}$, defined from

$$\vec{\mathbf{S}} = \frac{1}{\mu_0}\vec{\mathbf{E}}\times\vec{\mathbf{B}}. \qquad (38\text{-}20)$$

Let us first convince ourselves that $\vec{\mathbf{S}}$ as defined in Eq. 38-20 gives the direction of travel of the wave. According to the usual rule for the cross product, $\vec{\mathbf{S}}$ must be perpendicular to the plane determined by $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ in a direction determined by the right-hand rule. Applying the right-hand rule to Fig. 38-8, you can see that in the first half-cycle of the wave, where $\vec{\mathbf{E}}$ is in the $+y$ direction and $\vec{\mathbf{B}}$ is in the $+z$ direction, the cross product of $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ points in the $+x$ direction, which is indeed the direction of travel of the wave. In the second half-cycle, where $\vec{\mathbf{E}}$ is in the $-y$ direction and $\vec{\mathbf{B}}$ is in the $-z$ direction, the cross product is again in the $+x$ direction.

Note that an electromagnetic wave can be uniquely specified by giving only its $\vec{\mathbf{E}}$ field and its direction of travel. That is, using Eqs. 38-15 and 38-20 we can find the $\vec{\mathbf{B}}$ field of a wave if we know its $\vec{\mathbf{E}}$ field and its direction of travel. Equation 38-15 gives the magnitude of $\vec{\mathbf{B}}$ and Eq. 38-20 gives its direction.

Now let us examine the magnitude of $\vec{\mathbf{S}}$ as defined in Eq. 38-20 and demonstrate that it gives the power per unit area of the wave. For the wave shown in Fig. 38-8, the magnitude of $\vec{\mathbf{S}}$ is

$$S = \frac{1}{\mu_0}EB, \qquad (38\text{-}21)$$

---

* The Poynting vector is named for John Henry Poynting (1852–1914), who first discussed its properties. Poynting was a British physicist who was known for his studies of electromagnetism and gravitation.

where S, E, and B are the instantaneous values at the observation point. Using Eqs. 38-16 and 38-19, we can write this as

$$S = \frac{1}{\mu_0 c} E^2 = \epsilon_0 c E^2 \quad \text{or} \quad S = \frac{c}{\mu_0} B^2. \qquad (38\text{-}22)$$

We have previously obtained the energy density (energy per unit volume) at any point where an electric or magnetic field is present: $u_E = \frac{1}{2}\epsilon_0 E^2$ (Eq. 30-28) and $u_B = \frac{1}{2}B^2/\mu_0$ (Eq. 36-22). These equations, which were derived for static fields, apply equally well to time-varying fields. Where there is both an electric and a magnetic field (as in an electromagnetic wave), the total energy density at any point is

$$u = u_E + u_B = \frac{1}{2}\epsilon_0 E^2 + \frac{1}{2}\frac{B^2}{\mu_0} = \frac{S}{2c} + \frac{S}{2c} = \frac{S}{c}, \qquad (38\text{-}23)$$

where we have used Eqs. 38-22 to express E^2 and B^2 in terms of S. Note in Eq. 38-23 that $u_E = u_B$ everywhere the wave travels.

Figure 38-11 represents the instantaneous fields as the wave passes through a small volume of space surrounding point P. The volume has a thickness dx in the direction of travel of the wave and an area A transverse to its direction of travel. The volume dV can be expressed as $A\ dx$. The wave travels the distance dx in a time $dt = dx/c$, where c is the wave speed. The electromagnetic energy dU in the volume element dV is

$$dU = u\,dV = \frac{S}{c}\,dV = \frac{S}{c}A\,dx = SA\,dt. \qquad (38\text{-}24)$$

The magnitude of the Poynting vector is then

$$S = \frac{1}{A}\frac{dU}{dt} = \frac{P}{A}, \qquad (38\text{-}25)$$

where the rate of energy flow dU/dt has been set equal to the power P. Equation 38-25 shows that the magnitude of the Poynting vector gives the rate of energy flow or power per unit area of the wave. It is apparent from Eq. 38-25 that the SI unit for S is watts/meter^2.

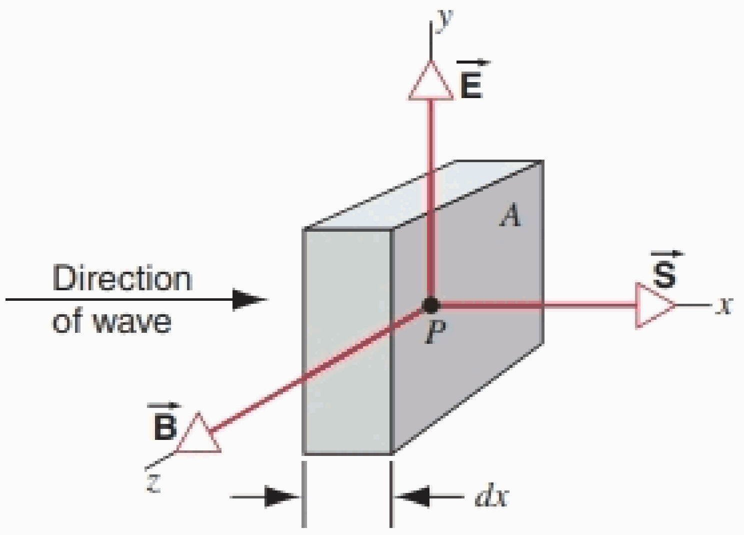


FIGURE 38-11. An electromagnetic wave passes through a small volume at point P. The fields and the Poynting vector at a particular instant of time are shown.

Intensity of an Electromagnetic Wave

Equation 38-21 relates the magnitude of S at any location to the magnitudes of E and B at that location at a particular instant of time. These values fluctuate very rapidly with time; for example, the frequency of a light wave is about 10^{15} Hz. For most detectors (our eyes, for example), this fluctuation is far too rapid to observe. Instead, we observe the *time average* of S, taken over many cycles of the wave. The time average S_{av} is also known as the *intensity* I of the wave.

From Eqs. 38-11 and 38-22 we have

$$I = S_{\text{av}} = \frac{1}{\mu_0 c}(E^2)_{\text{av}} = \frac{1}{\mu_0 c}E_{\text{m}}^2\,[\sin^2(kx - \omega t)]_{\text{av}}.$$

Usually, the time interval over which we average is much larger than the period of one cycle. In this case we can neglect the effect of fractions of a cycle and replace the $\sin^2$ by its average over any whole number of cycles, or $\frac{1}{2}$. We thus obtain

$$I = \frac{1}{2\mu_0 c}E_{\text{m}}^2 = \frac{1}{2\mu_0}E_{\text{m}}B_{\text{m}}. \qquad (38\text{-}26)$$

The intensity may also be expressed in terms of the rms (root-mean-square) magnitudes of the fields. With $E_{\text{m}} = \sqrt{2}E_{\text{rms}}$, we obtain

$$I = \frac{1}{\mu_0 c}E_{\text{rms}}^2 = \frac{1}{\mu_0}E_{\text{rms}}B_{\text{rms}}. \qquad (38\text{-}27)$$

Note that Eq. 38-27 looks like Eq. 38-21 if we replace the instantaneous values by their rms average values.

Figure 38-8 and Eqs. 38-11 and 38-12 represent a wave whose amplitudes E_{m} and B_{m} do not change with location as the wave travels. Light from a laser is a good approximation of such a wave. Another example is light from a very distant source that we observe over distances much smaller than the distance to the source—for example, light from the Sun. Often, however, we are dealing with nearby sources, for which Eqs. 38-11 and 38-12 with constant amplitudes are not valid. If we consider a point source of waves, such as a lightbulb viewed at distances much larger than the size of the bulb, the wavefronts from the source spread out as spheres if the source emits its radiation with equal intensity in all directions (called *isotropically*).

If the waves do not lose energy as they travel, then the energy in any spherical wavefront remains constant. That is, the rate at which energy from the wave passes through any spherical surface centered on the source is independent of the radius r of the sphere. As r increases, the total power delivered by the wavefront remains the same, but the power per unit area decreases, because the surface area of the sphere increases. If P is the average power emitted by the source, then the intensity on the spherical surface is the power per unit area:

$$I = \frac{P}{4\pi r^2}. \qquad (38\text{-}28)$$

The intensity of the wave from a point source decreases with distance from the source like $1/r^2$. Comparison of Eqs. 38-27 and 38-28 shows that the corresponding electric and magnetic field amplitudes (E_m and B_m or E_{rms} and B_{rms}) decrease like $1/r$.

SAMPLE PROBLEM 38-3. An observer is 1.8 m from a light source (of dimensions much smaller than 1.8 m) whose power output P is 250 W. Calculate the rms values of the electric and magnetic fields at the position of the observer. Assume that the source radiates uniformly in all directions.

Solution Combining Eqs. 38-27 and 38-28, we obtain

$$I = \frac{P}{4\pi r^2} = \frac{1}{\mu_0 c} E^2_{\rm rms}.$$

The rms electric field is

$$\begin{aligned} E_{\rm rms} &= \sqrt{\frac{P\mu_0 c}{4\pi r^2}} \\ &= \sqrt{\frac{(250\ \text{W})(4\pi \times 10^{-7}\ \text{H/m})(3.00 \times 10^8\ \text{m/s})}{(4\pi)(1.8\ \text{m})^2}} \\ &= 48\ \text{V/m}. \end{aligned}$$

The rms value of the magnetic field follows from Eq. 38-15 and is

$$\begin{aligned} B_{\rm rms} &= \frac{E_{\rm rms}}{c} = \frac{48\ \text{V/m}}{3.00 \times 10^8\ \text{m/s}} \\ &= 1.6 \times 10^{-7}\ \text{T} = 0.16\ \mu\text{T}. \end{aligned}$$

Note that $E_{\rm rms}$ (= 48 V/m) is appreciable as judged by ordinary laboratory standards but that $B_{\rm rms}$ (= 0.16 μT) is quite small. This helps to explain why most instruments used for the detection and measurements of electromagnetic waves respond to the electric component of the wave. It is wrong, however, to say that the electric component of an electromagnetic wave is "stronger" than the magnetic component, because we cannot compare quantities that are measured in different units. As we have seen, the electric and the magnetic components are on an absolutely equal basis as far as the propagation of the wave is concerned. Their average energies, which *can* be compared, are exactly equal.

38-7 RADIATION PRESSURE

When an electromagnetic wave is incident on an object, the object can absorb energy from the wave and the result is often observed as an increase in the object's temperature. In effect, the electric field of the wave exerts a force on the electrons to accelerate them. In collisions with atoms of the material, the electrons can transfer their energy to the entire object, thereby raising its temperature.

It is in principle possible for the absorption of an electromagnetic wave to transfer momentum to the entire object by exerting a force on its electrons. In effect, the wave exerts a net force on the object in the direction of travel of the wave. This force is generally very small—we do not feel the effects of this force when light shines on us—but it can be observed in the laboratory under carefully controlled conditions.

Figure 38-12 shows the electric and magnetic fields of an electromagnetic wave that is incident on a thin sheet of material of high resistivity. The electric field of the wave exerts a force $\vec{\mathbf{F}}_E = -e\vec{\mathbf{E}}$ on the electrons, which are accelerated in a direction opposite to the field (the $-y$ direction in Fig. 38-12). When an electric field is applied to a conducting material, we know that the electrons acquire a drift velocity $\vec{\mathbf{v}}_d$ that is proportional to the field strength (see, for example, Eq. 29-19). Ignoring all of the details of the properties of the material, let us write that the magnitude of the drift velocity is proportional to the electrical force, such that

$$bv_d = eE, \tag{38-29}$$

where b is a proportionality constant. This constant plays a similar role to the damping constant for a particle moving in a viscous fluid, such as a stone falling in a container of oil. If the effective damping force (that is, the resistivity) is large, then the speed of the electron will continually readjust to the electric field so that Eq. 38-29 remains valid even for the time-varying field of an electromagnetic wave.

As the electron moves along the y axis, the magnetic field of the wave exerts a force on it: $\vec{\mathbf{F}}_B = -e\vec{\mathbf{v}}_d \times \vec{\mathbf{B}}$. Figure 38-12 shows that this force is in the $+x$ direction—that is, the same direction that the wave is traveling. When $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ reverse direction, the force remains in the $+x$ direction. Because $\vec{\mathbf{v}}_d$ and $\vec{\mathbf{B}}$ are perpendicular to one another, we can write the magnitude of the magnetic force on a single electron as $F_B = ev_dB$. Substituting $v_d = eE/b$ from Eq. 38-29, we find

$$F_B = ev_dB = \frac{e^2EB}{b} = \frac{e^2E^2}{cb}, \tag{38-30}$$

where the last result follows from using $B = E/c$. If the sheet is uniformly illuminated and contains N electrons, the total force in the x direction is $F = NF_B$.

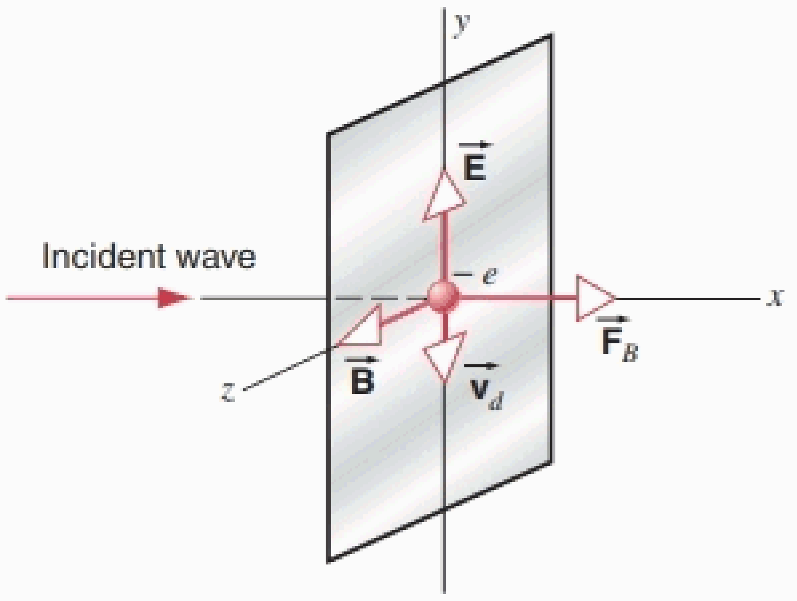


FIGURE 38-12. An incident plane light wave falls on an electron in a thin resistive sheet. Instantaneous values of $\vec{\mathbf{E}}$, $\vec{\mathbf{B}}$, the electron velocity $\vec{\mathbf{v}}_d$, and the radiation force $\vec{\mathbf{F}}_B$ are shown.

Let us also examine the rate at which the sheet absorbs energy from the wave. The $\vec{\mathbf{B}}$ field does no work on the electrons, because the magnetic force is always perpendicular to their motion, but the $\vec{\mathbf{E}}$ field does work and therefore transfers energy. The power (rate of energy transfer) delivered from the $\vec{\mathbf{E}}$ field to an electron is

$$\frac{dU_e}{dt} = F_E v_d = (eE)\left(\frac{eE}{b}\right) = \frac{e^2E^2}{b}. \qquad (38\text{-}31)$$

The total rate of energy absorption for all N electrons in the sheet is $dU/dt = N\,dU_e/dt$.

We can now obtain an expression for the total force on the sheet in terms of the rate of energy absorption by combining Eqs. 38-30 and 38-31:

$$F = NF_B = \frac{Ne^2E^2}{cb} = \frac{1}{c}\frac{dU}{dt} = \frac{SA}{c}, \qquad (38\text{-}32)$$

where A is the total area of the sheet. Taking the time average of all quantities, we find

$$\frac{F_{\text{av}}}{A} = \frac{I}{c} \quad \text{(absorption).} \qquad (38\text{-}33)$$

Equation 38-33 shows that the total pressure (force per unit area) exerted by the radiation on the sheet is proportional to the intensity I of the radiation.

Using Newton's second law in the form $F = dp/dt$, we can write Eq. 38-32 as

$$\frac{dp}{dt} = \frac{1}{c}\frac{dU}{dt} \quad \text{or} \quad dp = \frac{dU}{c}.$$

Integrating, we obtain

$$\Delta p = \frac{\Delta U}{c} \quad \text{(absorption).} \qquad (38\text{-}34)$$

This result gives the change in momentum Δp of an object that "recoils" after absorbing energy ΔU from an electromagnetic wave.

It is also possible for the object to reflect the radiation incident on it instead of absorbing it. Because the wave must reverse direction upon reflection, the object recoils with twice the momentum it would have in the case of absorption. (This case is similar to the gas molecule that imparts a pressure to the walls of a container in collisions that reverse its momentum; see Section 22-2.) In the case of reflection, the pressure exerted by the wave on the object and the momentum imparted to the object are both twice as large as in the case of absorption:

$$\frac{F_{\text{av}}}{A} = \frac{2I}{c} \quad \text{(reflection),} \qquad (38\text{-}35)$$

$$\Delta p = \frac{2\Delta U}{c} \quad \text{(reflection).} \qquad (38\text{-}36)$$

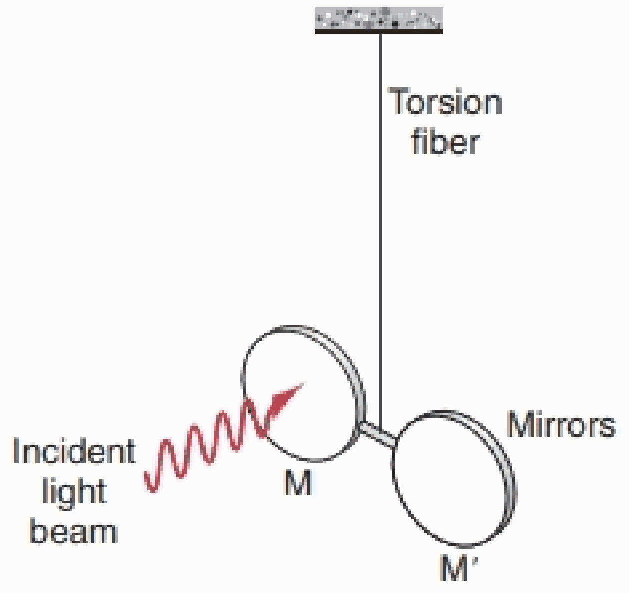


Figure 38-13. The arrangement of Nichols and Hull for measuring radiation pressure. The pressure of light on mirror M causes the fiber to twist through a small angle. Many details of this delicate experiment are omitted from the drawing.

The first measurements of radiation pressure were made in 1903 by E. F. Nichols and G. F. Hull. They allowed a beam of light to fall on a small mirror attached to a torsion fiber, as in Fig. 38-13. The pressure on the mirror caused the fiber to twist through a small angle θ. The pressure they measured was in the range 10^{-5} N/m^2, about 10^{-10} times smaller than atmospheric pressure. You can imagine the extraordinary care that the experimenters must have taken in order to observe this small effect. Today, lasers allow us to achieve much greater intensities of light, and the pressure of a laser beam can levitate a small object (Fig. 38-14).

Although the effects of radiation pressure from light falling on ordinary objects are difficult to observe, microscopic objects can be substantially affected by radiation pressure. Small grains of dust released by comets experience the radiation pressure from sunlight. This pressure pushes the grains of dust away from the Sun, where we

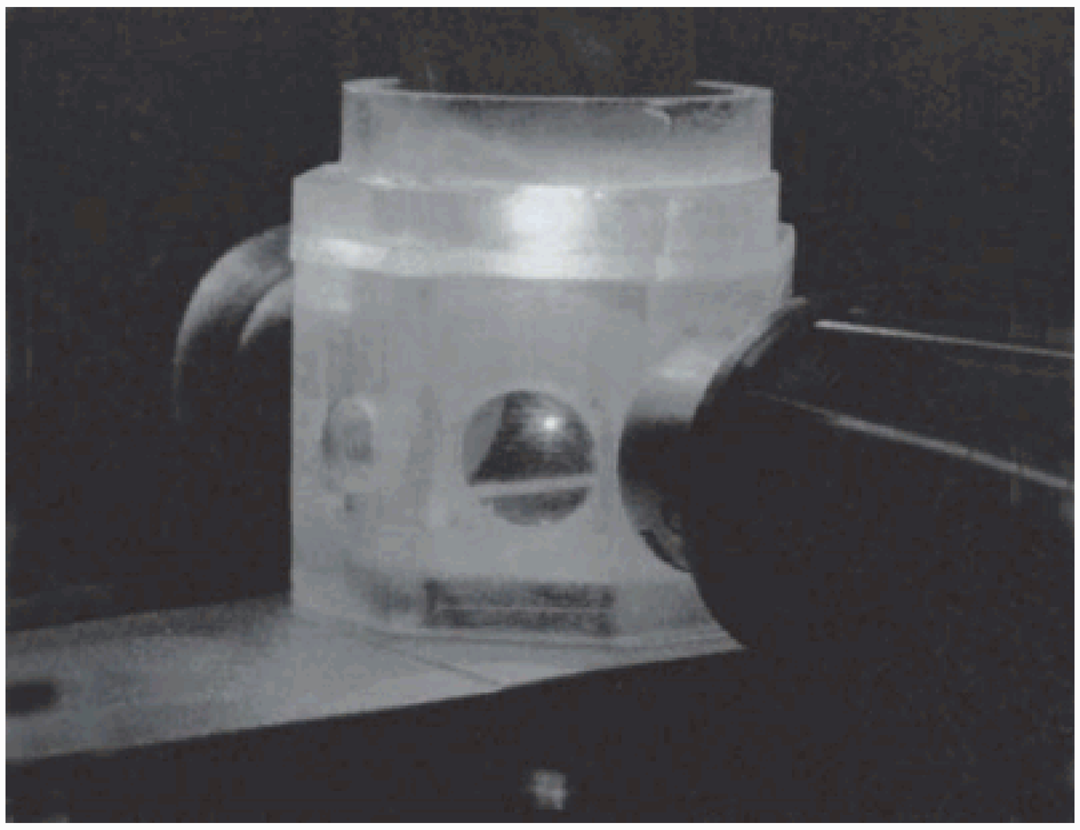


Figure 38-14. A small droplet of liquid (less than 10 μm in diameter) is levitated by a vertical laser beam. The horizontal tubes are microscopes used to observe the light scattered by the droplet. Courtesy David W. DuBois.

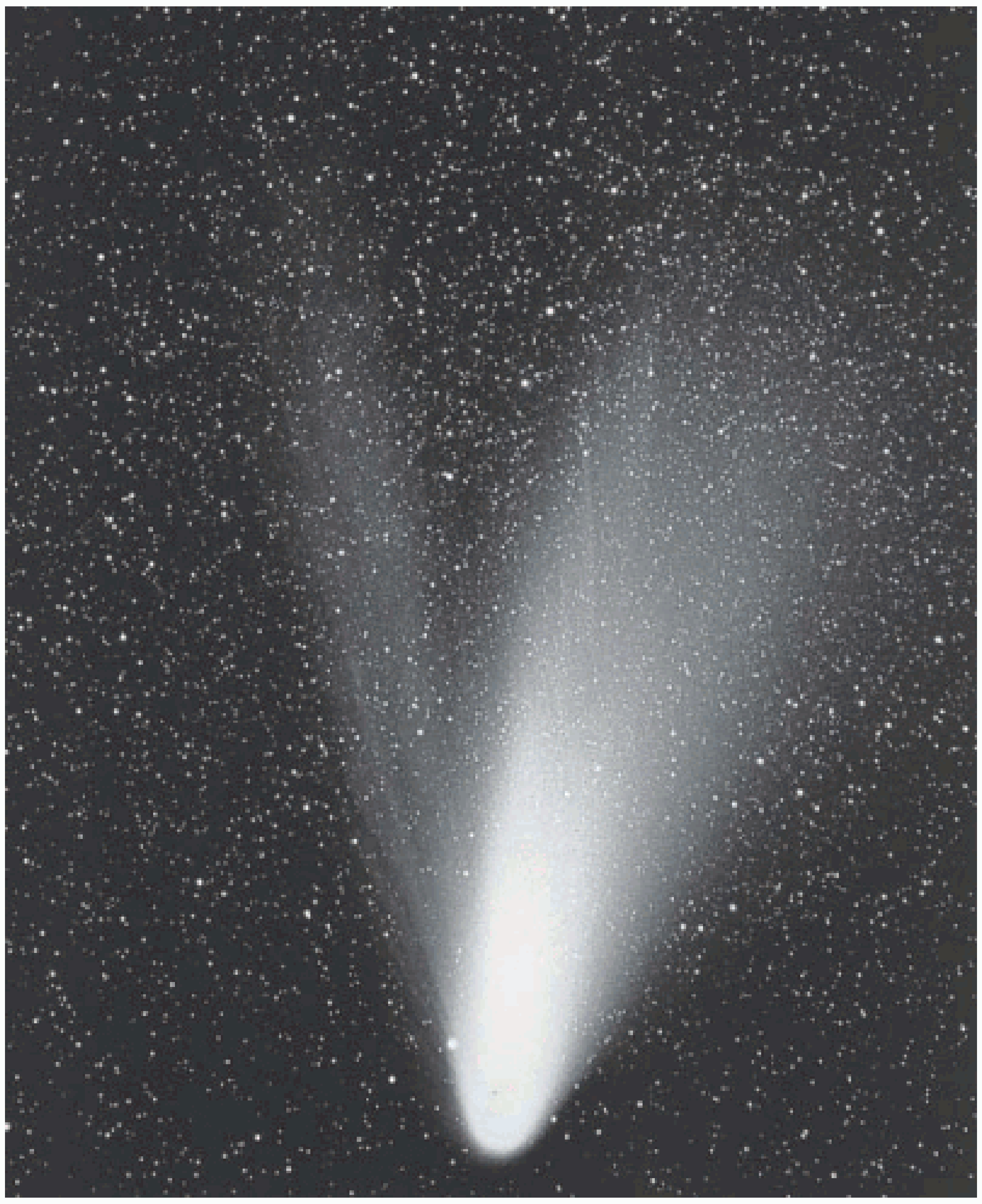


FIGURE 38-15. Comet Hale–Bopp, showing two tails. The brighter tail on the right is due to dust particles from the comet that are pushed away from the Sun by the radiation pressure. The fainter tail on the left is due to ions from the comet accelerated by the solar wind (high-speed protons and electrons from the Sun).

observe them as the "tail" of the comet (Fig. 38-15). It has even been proposed to build giant sheets of thin reflective material in space and use them to "sail" spacecraft through the solar system.

SAMPLE PROBLEM 38-4. Consider a dust particle released from a comet that is a distance R from the Sun and far from any planet. Assuming that the particle absorbs all the light incident on it, find the radius r of the particle if the force from the radiation pressure exactly balances the gravitational attraction of the Sun. Assume that the particle is spherical and has the same density as material of the Earth's crust, about $3.0 \times 10^3\ \text{kg/m}^3$.

Solution At first it may seem that we are not given enough information to solve this problem, since we do not know the distance R from the Sun. However, the gravitational force and the radiation intensity (Eq. 38-28) both vary like R^{-2}, and so the distance from the Sun should not affect our calculation.

The force exerted on the particle by the light pressure can be found from Eq. 38-33. For the area A over which the light is absorbed we use the circular area πr^2 of the dust particle. Setting the force due to the light pressure equal to the gravitational force, we obtain

$$\frac{IA}{c} = \frac{GmM}{R^2},$$

where m $(= \rho V)$ is the mass of the particle and M is the mass of the Sun. Substituting Eq. 38-28 for the radiation intensity, we have

$$\frac{P(\pi r^2)}{4\pi R^2 c} = \frac{GmM}{R^2} = \frac{G\rho(\frac{4}{3}\pi r^3)M}{R^2}.$$

Solving for r, we obtain

$$r = \frac{3P}{16\pi c G\rho M}.$$

Values of the radiant power of the Sun and its mass can be found in Appendix C. Substituting the numerical values, we find

$$r = 1.9 \times 10^{-7}\ \text{m} = 0.19\ \mu\text{m}.$$

As dust particles of this size leave the comet, no net force acts on them, and they continue to move in a straight line with velocity equal to that of the comet at the time of their release. Because the gravitational force grows with the mass of the particle (that is, proportional to r^3) while the radiation force grows with its cross-sectional area (that is, like r^2), larger particles fall toward the Sun but smaller particles are pushed away from the Sun. The net pattern of all such particles forms the tail of the comet.

MULTIPLE CHOICE

38-1 The Basic Equations of Electromagnetism

1. Where does the negative sign in Eq. 38-3 come from?

(A) Gauss' law (B) Faraday's law
(C) Lenz' law (D) Cole's law

38-2 Induced Magnetic Fields and the Displacement Current

2. The current through a long, straight wire is increasing according to $i = (4.0\ \mu\text{A/s})t$.

(*a*) What is the magnitude of the displacement current in the wire?

(A) $i_d = 4.0\ \mu\text{A}$
(B) $i_d = i = (4.0\ \mu\text{A/s})t$
(C) There is no displacement current in the wire.
(D) More information is needed to answer the question.

(*b*) What is the direction of the displacement current?

(A) Parallel to the original current
(B) Antiparallel to the original current
(C) There is no displacement current.

38-3 Maxwell's Equations

3. Which of Maxwell's equations in Table 38-1 are derived from purely theoretical arguments, requiring no measurements?

(A) All four equations
(B) II only
(C) I and II, the Gauss' law equations
(D) None of them

4. What term would need to be added to Faraday's law of induction if magnetic monopoles q_m were discovered? Assume that magnetic current is defined as $i_m = dq_m/dt$.

(A) $-\epsilon_0 i_m$ (B) $-i_m/\epsilon_0$
(C) $-i_m/\mu_0$ (D) $-\mu_0 i_m$

38-4 Generating an Electromagnetic Wave

5. An oscillating current in a wire loop lying in a horizontal plane will radiate electromagnetic waves. What is the direction of the electric field vector in the region far to the north of this loop?

(A) North/south (B) East/west (C) Up/down

6. AM radio waves are generated with a vertical electric dipole antenna; they are received with a long, cylindrical solenoid. How should the solenoid be oriented to best "pick-up" the AM signal?

(A) The axis of the solenoid should be vertical.
(B) The axis of the solenoid should be horizontal, and point toward the broadcasting antenna.
(C) The axis of the solenoid should be horizontal and perpendicular to the direction of the broadcasting antenna.
(D) It does not matter how the solenoid is oriented, because AM signals are isotropic.

38-5 Traveling Waves and Maxwell's Equations

7. Equation 38-6, $E = cB$, relates the

(A) instantaneous (B) average
(C) root-mean-square (D) maximum
(E) all of the above

values for E and B.

38-6 Energy Transport and the Poynting Vector

8. At a certain point and a certain time the electric field of an electromagnetic wave points north when the magnetic field points up. In which direction is the electromagnetic wave traveling?

(A) East (B) South
(C) West (D) Down

9. Of the three vectors in the equation $\vec{\mathbf{S}} = \frac{1}{\mu_0}\vec{\mathbf{E}} \times \vec{\mathbf{B}}$, which pair(s) are always at right angles for a plane electromagnetic wave in empty space? *(There may be more than one correct answer.)*

(A) $\vec{\mathbf{S}}$ and $\vec{\mathbf{E}}$ (B) $\vec{\mathbf{S}}$ and $\vec{\mathbf{B}}$
(C) $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ (D) None
(E) All three must be at right angles.

10. An electromagnetic radiation source radiates uniformly in all directions. How does the magnitude of the electric field strength vary with distance r from the source?

(A) E_m is constant for electromagnetic waves.
(B) $E_m \propto 1/r$
(C) $E_m \propto 1/r^2$
(D) $E_m \propto 1/r^3$

38-7 Radiation Pressure

11. An electromagnetic wave is allowed to reflect from a mirror in a measurement of radiation pressure. If E_{rms} is doubled while the wave frequency is cut in half, then the radiation pressure will

(A) quadruple. (B) double.
(C) remain the same. (D) be cut in half.

QUESTIONS

1. In your own words, explain why Faraday's law of induction (see Table 38-1) can be interpreted by saying "a changing magnetic field generates an electric field."

2. If a uniform flux Φ_E through a plane circular ring decreases with time, is the induced magnetic field (as viewed along the direction of $\vec{\mathbf{E}}$) clockwise or counterclockwise?

3. If (as is true) there are unit systems in which ϵ_0 and μ_0 do not appear, how can Eq. 38-19 be true?

4. Why is it so easy to show that "a changing magnetic field produces an electric field" but so hard to show in a simple way that "a changing electric field produces a magnetic field"?

5. In Fig. 38-2, consider a circle with $r > R$. How can a magnetic field be induced around this circle, as Sample Problem 38-1 shows? After all, there is no electric field at the location of this circle, and $dE/dt = 0$ here.

6. In Fig. 38-2, $\vec{\mathbf{E}}$ is into the figure and is increasing in magnitude. Find the direction of $\vec{\mathbf{B}}$ if, instead, (*a*) $\vec{\mathbf{E}}$ is into the figure and decreasing, (*b*) $\vec{\mathbf{E}}$ is out of the figure and increasing, (*c*) $\vec{\mathbf{E}}$ is out of the figure and decreasing, and (*d*) $\vec{\mathbf{E}}$ remains constant.

7. In Fig. 36-8*c*, a displacement current is needed to maintain continuity of current in the capacitor. How can one exist, considering that there is no charge on the capacitor?

8. (*a*) In Fig. 38-2, what is the direction of the displacement current i_d? In this same figure, can you find a rule relating the directions (*b*) of $\vec{\mathbf{B}}$ and $\vec{\mathbf{E}}$ and (*c*) of $\vec{\mathbf{B}}$ and $d\vec{\mathbf{E}}/dt$?

9. What advantages are there in calling the term $\epsilon_0 d\Phi_E/dt$ in Eq. IV, Table 38-1, a displacement current?

10. Can a displacement current be measured with an ammeter? Explain.

11. Why are the magnetic fields of conduction currents in wires so easy to detect but the magnetic effects of displacement current in capacitors so hard to detect?

12. In Table 38-1 there are three kinds of apparent lack of symmetry in Maxwell's equations. (*a*) The quantities ϵ_0 and/or μ_0 appear in I and IV but not in II and III. (*b*) There is a negative sign in III but no negative sign in IV. (*c*) There are missing "magnetic pole terms" in II and III. Which of these represent genuine lack of symmetry? If magnetic monopoles were discovered, how would you rewrite these equations to include them? (Hint: Let q_m be the magnetic pole strength, analogous to the quantum of charge e; what SI units would q_m have?)

13. Maxwell's equations as displayed in Table 38-1 are written on the assumption that no dielectric materials are present. How should the equations be written if this restriction is removed?

14. Discuss the periodic flow of energy, if any, from point to point in an acoustic resonant cavity.

15. An air-filled acoustic resonant cavity and an electromagnetic resonant cavity of the same size have resonant frequencies that are in the ratio of 10^6 or so. Which has the higher frequency and why?

16. Electromagnetic cavities are often silver-plated on the inside. Why?

17. At what parts of the cycle will (*a*) the conduction current and (*b*) the displacement current in the cavity of Fig. 38-3 be zero?

18. Discuss the time variation during one complete cycle of the charges that appear at various points on the inner walls of the oscillating electromagnetic cavity of Fig. 38-3.

19. Speaking loosely, we can say that the electric and the magnetic components of a traveling electromagnetic wave "feed on each other." What does this mean?

20. "Displacement currents are present in a traveling electromagnetic wave and we may associate the magnetic field component of the wave with these currents." Is this statement true? Discuss it in detail.

21. Can an electromagnetic wave be deflected by a magnetic field? By an electric field?

22. Why is Maxwell's modification of Ampère's law (that is, the term $\mu_0\epsilon_0 d\Phi_E/dt$ in Table 38-1) needed to understand the propagation of electromagnetic waves?

23. Is it conceivable that electromagnetic theory might some day be able to predict the value of c (3×10^8 m/s), not in terms of μ_0 and ϵ_0, but directly and numerically without recourse to any measurement?

24. If you were to calculate the Poynting vector for various points in and around a transformer, what would you expect the field pattern to look like? Assume that an alternating potential difference has been applied to the primary windings and that a resistive load is connected across the secondary windings.

25. Name two historic experiments, in addition to the radiation pressure measurements of Nichols and Hull, in which a torsion balance was used. Both are described in this book—one in Volume 1 and one in Volume 2.

26. Can an object absorb light energy without having linear momentum transferred to it? If so, give an example. If not, explain why.

27. When you turn on a flashlight, does it experience any force associated with the emission of the light?

28. We associated energy and linear momentum with electromagnetic waves. Is angular momentum present also?

29. What is the relation, if any, between the intensity I of an electromagnetic wave and the magnitude S of its Poynting vector?

30. As you recline in a beach chair in the Sun, why are you so conscious of the thermal energy delivered to you but totally unresponsive to the linear momentum delivered from the same source? Is it true that when you catch a hard-pitched baseball, you are conscious of the energy delivered but not of the momentum?

31. When a parallel beam of light falls on an object, the momentum transfers are given by Eqs. 38-34 and 38-36. Do these equations still hold if the light source is moving rapidly toward or away from the object at, perhaps, a speed of $0.1c$?

32. Radiation pressure is believed responsible for setting an upper limit (of about $100M_{\text{Sun}}$) to the mass of a star. Explain.

EXERCISES

38-1 The Basic Equations of Electromagnetism

38-2 Induced Magnetic Fields and the Displacement Current

1. For the situation of Sample Problem 38-1, where is the induced magnetic field equal to one-half of its maximum value?

2. Prove that the displacement current in a parallel-plate capacitor can be written

$$i_d = C\frac{dV}{dt}.$$

3. You are given a 1.0-μF parallel-plate capacitor. How would you establish an (instantaneous) displacement current of 1.0 mA in the space between its plates?

4. In Sample Problem 38-1, show that the *displacement current density* j_d is given, for $r < R$, by

$$j_d = \epsilon_0\frac{dE}{dt}.$$

5. A parallel-plate capacitor has square plates 1.22 m on a side as in Fig. 38-16. There is a charging current of 1.84 A flowing into (and out of) the capacitor. (*a*) What is the displacement current through the region between the plates? (*b*) What is dE/dt in this region? (*c*) What is the displacement current through the square dashed path between the plates? (*d*) What is $\oint \vec{B}\cdot d\vec{s}$ around this square dashed path?

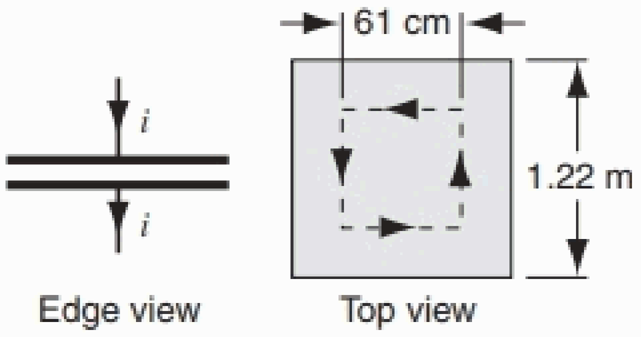


FIGURE 38-16. Exercise 5.

6. In Sample Problem 38-1, show that the expressions derived for $B(r)$ can be written

$$B(r) = \frac{\mu_0 i_d}{2\pi r} \qquad (r \geq R),$$

$$B(r) = \frac{\mu_0 i_d r}{2\pi R^2} \qquad (r \leq R).$$

Note that these expressions are of the same form as those derived in Chapter 33 except that the conduction current i has been replaced by the displacement current i_d.

7. A uniform electric field collapses to zero from an initial strength of 0.60 MV/m in a time of 15 μs in the manner shown in Fig. 38-17. Calculate the displacement current, through a 1.9 m^2 region perpendicular to the field, during each of the time intervals (*a*), (*b*), and (*c*) shown on the graph. (Ignore the behavior at the ends of the intervals.)

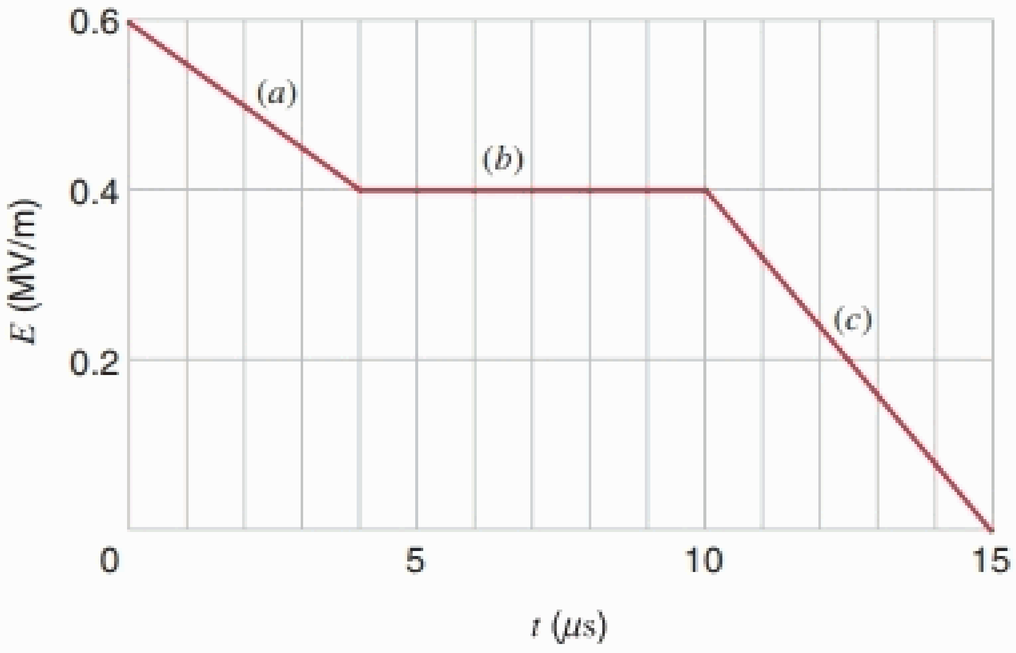


FIGURE 38-17. Exercise 7.

8. A parallel-plate capacitor with circular plates 21.6 cm in diameter is being charged as in Fig. 38-2. The displacement current density throughout the region is uniform, into the paper in the diagram, and has a value of 1.87 mA/cm^2. (*a*) Calculate the magnetic field B at a distance r = 53.0 mm from the axis of symmetry of the region. (*b*) Calculate dE/dt in this region.

9. Suppose that a circular-plate capacitor has a radius R of 32.1 mm and a plate separation of 4.80 mm. A sinusoidal potential difference with a maximum value of 162 V and a frequency of 60.0 Hz is applied between the plates. Find the maximum value of the induced magnetic field at $r = R$.

38-3 Maxwell's Equations

10. Collect and tabulate expressions for the following four quantities, considering both $r < R$ and $r > R$. Place the derivations side by side and study them as interesting applications of Maxwell's equations to problems having cylindrical symmetry. (*a*) $B(r)$ for a current i in a long wire of radius R. (*b*) $E(r)$ for a long uniform cylinder of charge of radius R. (*c*) $B(r)$ for a parallel-plate capacitor, with circular plates of radius R, in which E is changing at a constant rate. (*d*) $E(r)$ for a cylindrical region of radius R in which a uniform magnetic field B is changing at a constant rate.

11. Two adjacent closed paths *abefa* and *bcdeb* share the common edge *be* as shown in Fig. 38-18. (*a*) We may apply $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -\,d\Phi_B/dt$ (Eq. III of Table 38-1) to each of these two closed paths separately. Show that, from this alone, Eq. III is automatically satisfied for the composite path *abcdefa*. (*b*) Repeat using Eq. IV. (*c*) This relation is called a "self-consistency" property; why must each of Maxwell's equations be self-consistent?

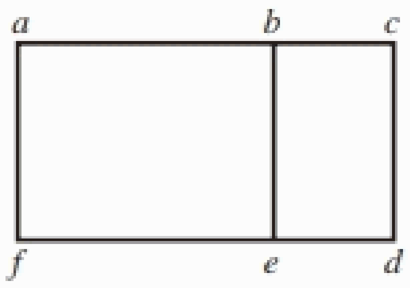


FIGURE 38-18. Exercise 11.

12. Two adjacent closed parallelepipeds share a common face as shown in Fig. 38-19. (*a*) We may apply $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q/\epsilon_0$ (Eq. I in Table 38-1) to each of these two closed surfaces separately. Show that, from this alone, Eq. I is automatically satisfied for the composite closed surface. (*b*) Repeat using Eq. II. See Exercise 11.

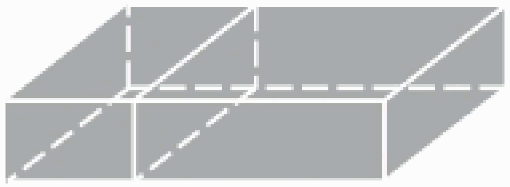


FIGURE 38-19. Exercise 12.

13. In microscopic terms the principle of continuity of current may be expressed as

$$\oint (\vec{\mathbf{j}} + \vec{\mathbf{j}}_{\mathrm{d}}) \cdot d\vec{\mathbf{A}} = 0,$$

in which $\vec{\mathbf{j}}$ is the conduction current density and $\vec{\mathbf{j}}_{\mathrm{d}}$ is the displacement current density. The integral is to be taken over any closed surface; the equation essentially says that whatever current flows into the enclosed volume must also flow out. Apply this equation to the surface shown by the dashed lines in Fig. 38-20 shortly after switch S is closed.

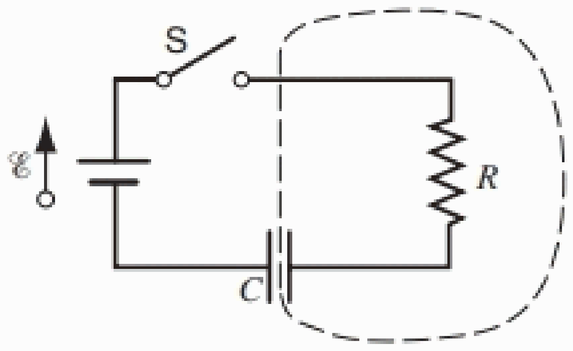


FIGURE 38-20. Exercise 13.

38-4 Generating an Electromagnetic Wave

14. Figure 38-21 shows an LC oscillator connected by a transmission line to an antenna of a magnetic dipole type. Compare with Fig. 38-5, which shows a similar arrangement but with an electric dipole type of antenna. (*a*) What is the basis for the names of these two antenna types? (*b*) Draw figures corresponding to Figs. 38-6 and 38-7 to describe the electromagnetic wave that sweeps past the observer at point P in Fig. 38-21.

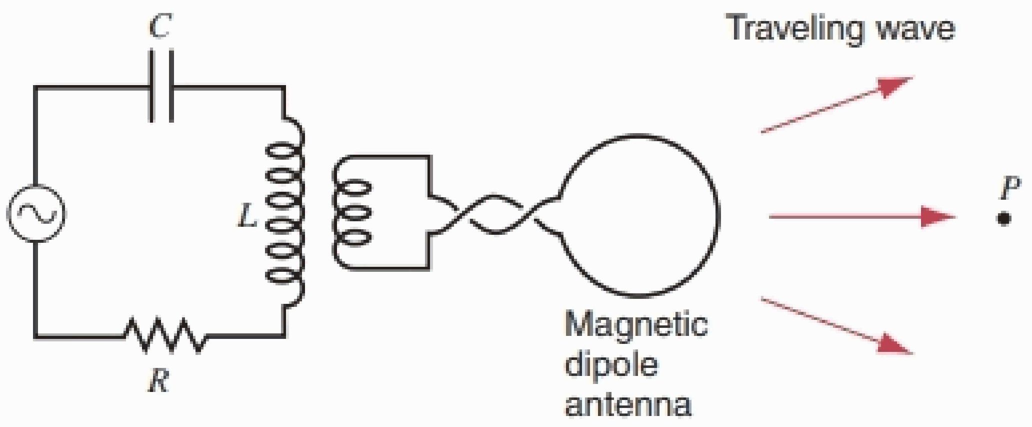


FIGURE 38-21. Exercise 14.

15. What inductance is required with a 17-pF capacitor in order to construct an oscillator capable of generating 550-nm (i.e., visible) electromagnetic waves? Comment on your answer.

38-5 Traveling Waves and Maxwell's Equations

16. The electric field associated with a plane electromagnetic wave is given by $E_x = 0$, $E_y = 0$, $E_z = E_0 \sin k(x - ct)$, where $E_0 = 2.34 \times 10^{-4}$ V/m and $k = 9.72 \times 10^6\ \text{m}^{-1}$. The wave is propagating in the $+x$ direction. (*a*) Write expressions for the components of the magnetic field of the wave. (*b*) Find the wavelength of the wave.

17. A certain plane electromagnetic wave has a maximum electric field of 321 μV/m. Find the maximum magnetic field.

18. Start from Eqs. 38-14 and 38-17 and show that $E(x, t)$, and $B(x, t)$, the electric and magnetic field components of a plane traveling electromagnetic wave, must satisfy the "wave equations"

$$\frac{\partial^2 E}{\partial t^2} = c^2 \frac{\partial^2 E}{\partial x^2} \qquad \text{and} \qquad \frac{\partial^2 B}{\partial t^2} = c^2 \frac{\partial^2 B}{\partial x^2}.$$

19. (*a*) Show that Eqs. 38-11 and 38-12 satisfy the wave equations displayed in Exercise 18. (*b*) Show that any expressions of the form $E = E_m f(kx \pm \omega t)$ and $B = B_m f(kx \pm \omega t)$, where $f(kx \pm \omega t)$ denotes an arbitrary function, also satisfy these wave equations.

38-6 Energy Transport and the Poynting Vector

20. Show, by finding the direction of the Poynting vector $\vec{\mathbf{S}}$, that the directions of the electric and magnetic fields at all points in Figs. 38-6, 38-7, and 38-8 are consistent at all times with the assumed directions of propagation.

21. Currently operating neodymium–glass lasers can provide 100 TW of power in 1.0-ns pulses at a wavelength of 0.26 μm. How much energy is contained in a single pulse?

22. A plane electromagnetic wave is traveling in the negative y direction. At a particular position and time, the magnetic field is along the positive z axis and has a magnitude of 28 nT. What are the direction and magnitude of the electric field at that position and at that time?

23. Our closest stellar neighbor, α-Centauri, a 4.30 light-years away. It has been suggested that TV programs from our planet have reached this star and may have been viewed by the hypothetical inhabitants of a hypothetical planet orbiting this star. A TV station on Earth has a power output of 960 kW. Find the intensity of its signal at α-Centauri.

24. (*a*) Show that in a plane traveling electromagnetic wave the average intensity—that is, the average rate of energy transport per unit area—is given by

$$S_{\text{av}} = \frac{cB_m^2}{2\mu_0}.$$

(*b*) What is the average intensity of a plane traveling electromagnetic wave if B_m, the maximum value of its magnetic field component, is 1.0×10^{-4} T?

25. The intensity of direct solar radiation not absorbed by the atmosphere on a particular summer day is 130 W/m^2. How close would you have to stand to a 1.0-kW electric heater to feel the same intensity? Assume that the heater radiates uniformly in all directions.

26. Prove that, for any point in an electromagnetic wave such as that of Fig. 38-8, the density of the energy stored in the electric field equals that of the energy stored in the magnetic field.

27. You walk 162 m directly toward a street lamp and find that the intensity increases to 1.50 times the intensity at your original position. (*a*) How far from the lamp were you first standing? (The lamp radiates uniformly in all directions.) (*b*) Can you find the power output of the lamp? If not, explain why.

28. Sunlight strikes the Earth, just outside its atmosphere, with an intensity of 1.38 kW/m^2. Calculate (*a*) E_m and (*b*) B_m for sunlight, assuming it to be a plane wave.

29. The maximum electric field at a distance of 11.2 m from a point light source is 1.96 V/m. Calculate (*a*) the amplitude of the magnetic field, (*b*) the intensity, and (*c*) the power output of the source.

30. Frank D. Drake, an active investigator in the SETI (Search for Extra-Terrestrial Intelligence) program, has said that the large radio telescope in Arecibo, Puerto Rico, "can detect a signal which lays down on the entire surface of the Earth a power of only one picowatt." See Fig. 38-22. (*a*) What is the power actually received by the Arecibo antenna for such a signal? The antenna diameter is 305 m. (*b*) What would be the power output of a source at the center of our galaxy that could provide such a signal? The galactic center is 2.3×10^4 ly away. Take the source as radiating uniformly in all directions.

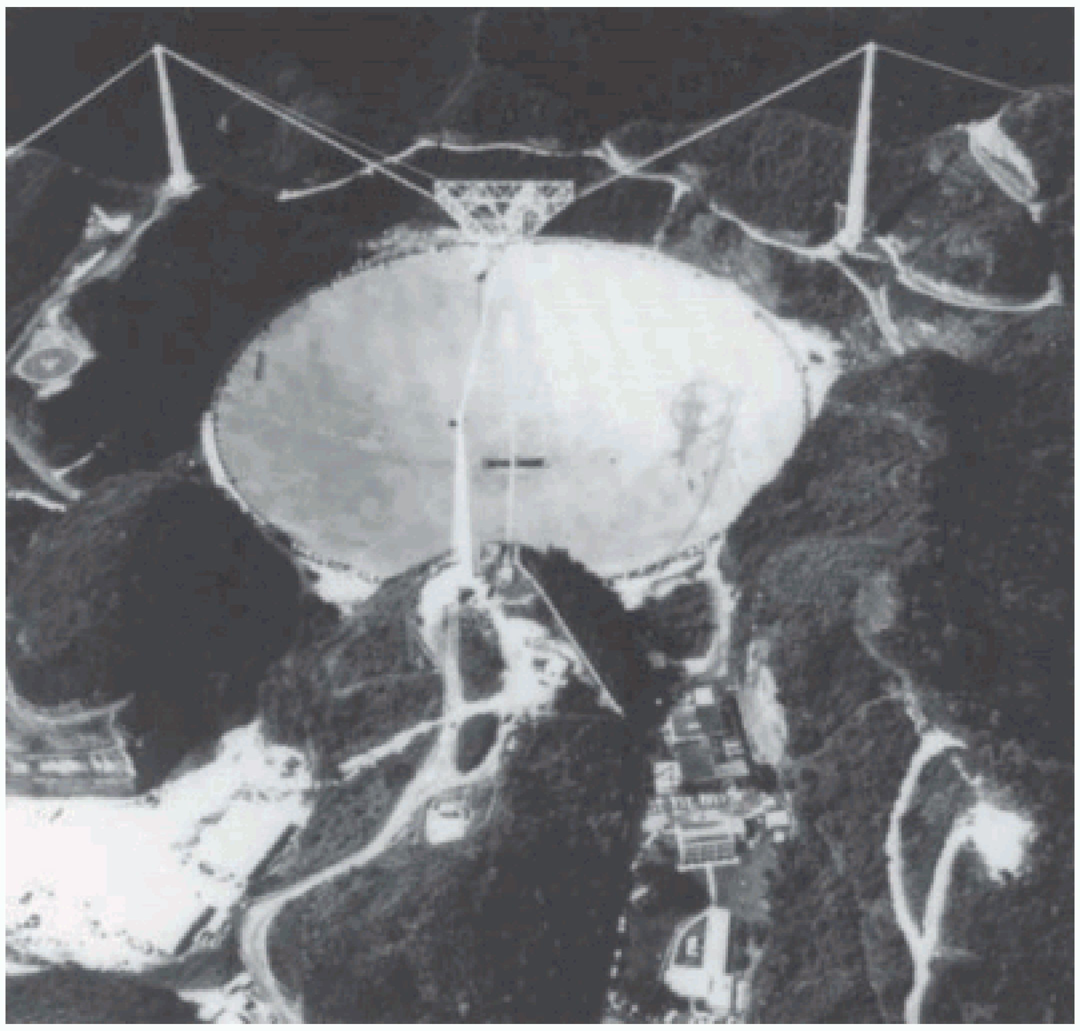


FIGURE 38-22. Exercise 30.

31. An airplane flying at a distance of 11.3 km from a radio transmitter receives a signal of 7.83 μW/m^2. Calculate (*a*) the amplitude of the electric field at the airplane due to this signal; (*b*) the amplitude of the magnetic field at the airplane; (*c*) the total power radiated by the transmitter, assuming the transmitter radiates uniformly in all directions.

32. During a test, a NATO surveillance radar system, operating at 12 GHz with 183 kW of output power, attempts to detect an incoming "enemy" aircraft at 88.2 km. The target aircraft is designed to have a very small effective area of 0.222 m^2 for reflection of radar waves. Assume that the radar beam spreads out isotropically into the forward hemisphere on both transmission and reflection, and ignore absorption in the atmosphere. For the reflected beam as received back at the radar site, calculate (*a*) the intensity, (*b*) the amplitude of the electric field vector, and (*c*) the rms value of the magnetic field.

38-7 Radiation Pressure

33. Suppose that you lie in the Sun for 2.5 h, exposing an area of 1.3 m^2 at 90° to the Sun's rays of intensity 1.1 kW/m^2. Assuming complete absorption of the rays, how much momentum is delivered to your body?

34. The average intensity of the solar radiation that falls normally on a surface just outside the Earth's atmosphere is 1.38 kW/m^2. (*a*) What radiation pressure is exerted on this surface, assuming complete absorption? (*b*) How does this pressure compare with the Earth's sea-level atmospheric pressure, which is 101 kPa?

35. High-power lasers are used to compress gas plasmas by radiation pressure. The reflectivity of a plasma is unity if the electron density is high enough. A laser generating pulses of radiation of peak power 1.5 GW is focused onto 1.3 mm^2 of high-electron-density plasma. Find the pressure exerted on the plasma.

36. Calculate the radiation pressure 1.50 m away from a 500-W lightbulb. Assume that the surface on which the pressure is exerted faces the bulb and is perfectly absorbing and that the bulb radiates uniformly in all directions.

37. Radiation from the Sun striking the Earth has an intensity of 1.38 kW/m^2. (*a*) Assuming that the Earth behaves like a flat disk at right angles to the Sun's rays and that all the incident energy is absorbed, calculate the force on the Earth due to radiation pressure. (*b*) Compare it with the force due to the Sun's gravitational attraction by calculating the ratio $F_{\text{rad}}/F_{\text{grav}}$.

38. Show that the vector $c\epsilon_0\vec{\mathbf{E}} \times \vec{\mathbf{B}}$ has the dimensions of momentum/(area · time), whereas $\mu_0^{-1}\ \vec{\mathbf{E}} \times \vec{\mathbf{B}}$ has the dimensions of energy/(area · time). (The vector $c\epsilon_0\vec{\mathbf{E}} \times \vec{\mathbf{B}}$ may be used for computing momentum flow in the same manner that $\vec{\mathbf{S}} = \mu_0^{-1}\ \vec{\mathbf{E}} \times \vec{\mathbf{B}}$ is used to compute energy flow.)

39. Radiation of intensity I is normally incident on an object that absorbs a fraction f of it and reflects the rest. What is the radiation pressure?

40. Prove, for a plane wave at normal incidence on a plane surface, that the radiation pressure on the surface is equal to the energy density in the beam outside the surface. This relation holds no matter what fraction of the incident energy is reflected.

41. Prove, for a stream of bullets striking a plane surface at normal incidence, that the "pressure" is twice the kinetic energy density in the stream above the surface; assume that the bullets are completely absorbed by the surface. Contrast this with the behavior of light; see Exercise 40.

42. A small spaceship whose mass, with occupant, is 1500 kg is drifting in outer space, where the gravitational field is negligible. If the astronaut turns on a 10.0-kW laser beam, what speed would the ship attain in one day because of the reaction force associated with the momentum carried away by the beam?

43. A laser has a power output of 4.6 W and a beam diameter of 2.6 mm. If it is aimed vertically upward, what is the height H of a perfectly reflecting cylinder that can be made to "hover" by the radiation pressure exerted by the beam? Assume that the density of the cylinder is 1.2 g/cm^3. See Fig. 38-23.

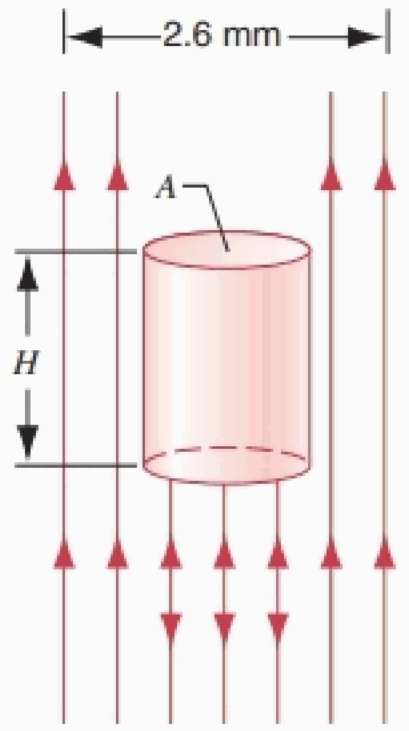


FIGURE 38-23. Exercise 43.

44. Exercise 8 of Chapter 3 described the Sun yacht *Diana*, designed to navigate the solar system using the pressure of sunlight. The sail area is 3.1 km^2, and the mass is 930 kg. Verify that the maximum radiation force on the sail is 29 N at a distance from the Sun equal to Earth's orbital radius.

PROBLEMS

1. Figure 38-24 shows the plates P_1 and P_2 of a circular parallel-plate capacitor of radius R. They are connected as shown to long, straight wires in which a constant conduction current i exists. Also shown are three hypothetical circles of radius r, two of them outside the capacitor and one between the plates. Show that the magnetic field at the circumference of each of these circles is given by

$$B = \frac{\mu_0 i}{2\pi r}.$$

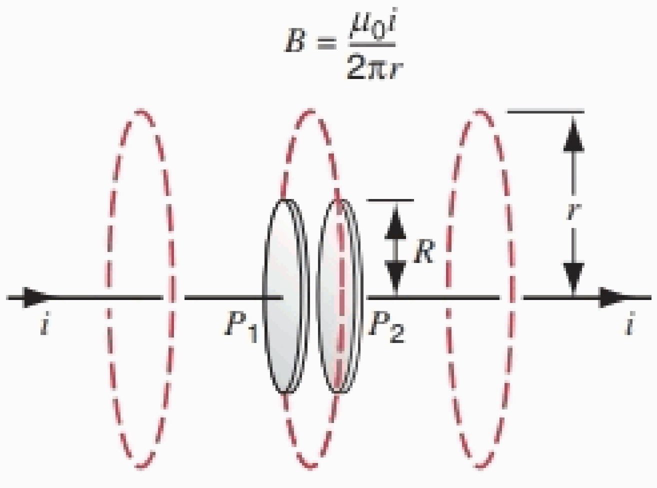


FIGURE 38-24. Problem 1.

2. In 1929, M. R. Van Cauwenberghe succeeded in measuring directly, for the first time, the displacement current i_{d} between the plates of a parallel-plate capacitor to which an alternating potential difference was applied, as suggested by Fig. 38-2. He used circular plates whose effective radius was 40.0 cm and whose capacitance was 100 pF. The applied potential difference had a maximum value ΔV_{m} of 174 kV at a frequency of 50.0 Hz. (*a*) What maximum displacement current was present between the plates? (*b*) Why was the applied potential difference chosen to be so high? (The delicacy of these measurements is such that they were only performed in a direct manner more than 60 years after Maxwell enunciated the concept of displacement current! The experiment is described in *Journal de Physique,* No. 8, 1929.)

3. The capacitor in Fig. 38-25 consisting of two circular plates with radius $R = 18.2$ cm is connected to a source of emf $\mathcal{E} = \mathcal{E}_{\mathrm{m}} \sin \omega t$, where $\mathcal{E}_{\mathrm{m}} = 225$ V and $\omega = 128$ rad/s. The maximum value of the displacement current is $i_{\mathrm{d}} = 7.63\ \mu$A. Neglect fringing of the electric field at the edges of the plates. (*a*) What is the maximum value of the current i? (*b*) What is the maximum value of $d\Phi_E/dt$, where Φ_E is the electric flux through the region between the plates? (*c*) What is the separation d between the plates? (*d*) Find the maximum value of the magnitude of $\vec{\mathbf{B}}$ between the plates at a distance $r = 11.0$ cm from the center.

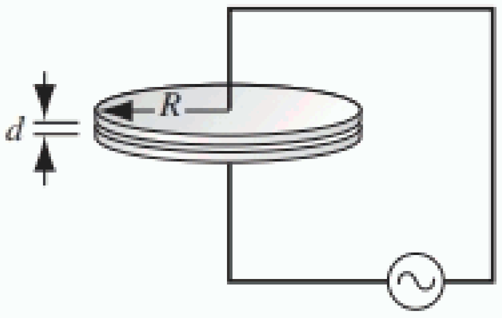


FIGURE 38-25. Problem 3.

4. A long, cylindrical conducting rod with radius R is centered on the x axis as shown in Fig. 38-26. A narrow saw cut is made in the rod at $x = b$. A conduction current i, increasing with time and given by $i = \alpha t$, flows toward the right in the rod; α is a (positive) proportionality constant. At $t = 0$ there is no charge on the cut faces near $x = b$. (*a*) Find the magnitude of the charge on these faces, as a function of time. (*b*) Use Eq. I in Table 38-1 to find E in the gap as a function of time. (*c*) Sketch the lines of $\vec{\mathbf{B}}$ for $r < R$, where r is the distance from the x axis. (*d*) Use Eq. IV in Table 38-1 to find $B(r)$ in the gap for $r < R$. (*e*) Compare the above answer with $B(r)$ in the rod for $r < R$.

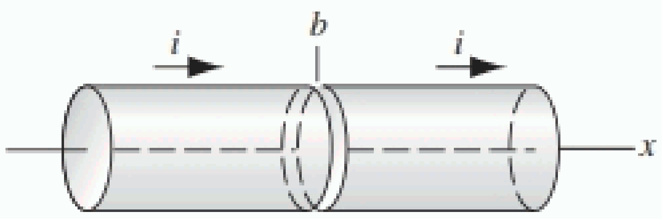


FIGURE 38-26. Problem 4.

5. A cube of edge a has its edges parallel to the x, y, and z axes of a rectangular coordinate system. A uniform electric field $\vec{\mathbf{E}}$ is parallel to the y axis and a uniform magnetic field $\vec{\mathbf{B}}$ is parallel to the x axis. Calculate (*a*) the rate at which, according to the Poynting vector point of view, energy may be said to pass through each face of the cube and (*b*) the net rate at which the energy stored in the cube may be said to change.

6. A cylindrical electromagnetic cavity 4.8 cm in diameter and 7.3 cm long is oscillating in the mode shown in Fig. 38-3. (*a*) Assume that, for points on the axis of the cavity, $E_{\mathrm{m}} =$ 13 kV/m. The frequency of oscillation is 2.4 GHz. For such axial points, what is the maximum rate $(dE/dt)_{\mathrm{m}}$, at which E changes? (*b*) Assume that the average value of $(dE/dt)_{\mathrm{m}}$, for all points over a cross section of the cavity, is one-half the value found above for axial points. On this assumption, what is the maximum value of B at the cylindrical surface of the cavity?

7. The average intensity of sunlight, falling at normal incidence just outside the Earth's atmosphere, varies during the year due to the changing Earth–Sun distance. Show that the fractional yearly variation is given by $\Delta I/I = 4e$ approximately, where e is the eccentricity of the Earth's elliptical orbit around the Sun.

8. The radiation emitted by a laser is not exactly a parallel beam; rather, the beam spreads out in the form of a cone with circular cross section. The angle θ of the cone (see Fig. 38-27) is called the full-angle beam divergence. A 3.85-kW argon laser, radiating at 514.5 nm, is aimed at the Moon in a ranging experiment; the laser has a full-angle beam divergence of 0.880 μrad. Find the intensity of the beam at the Moon's surface.

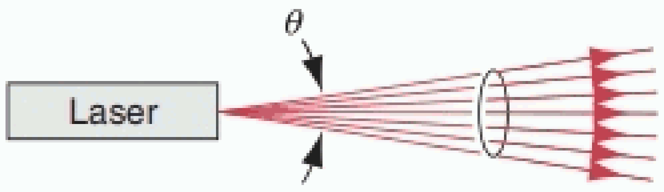


FIGURE 38-27. Problem 8.

9. Consider the possibility of standing electromagnetic waves:

$$E = E_{\mathrm{m}} \sin kx \sin \omega t,$$

$$B = B_{\mathrm{m}} \cos kx \cos \omega t.$$

(*a*) Show that these satisfy Eqs. 38-14 and 38-17 if E_{m} is suitably related to B_{m} and ω is suitably related to k. What are these relationships? (*b*) Find the (instantaneous) Poynting vector. (*c*) Show that the time-average power flow across any area is zero. (*d*) Describe the flow of energy in this situation.

10. A copper wire (diameter = 2.48 mm; resistance 1.00 Ω per 300 m) carries a current of 25.0 A. Calculate (*a*) the electric field, (*b*) the magnetic field, and (*c*) the Poynting vector magnitude for a point on the surface of the wire.

11. A coaxial cable (inner radius a, outer radius b) is used as a transmission line between a battery $\mathcal{E}$ and a resistor R, as shown in Fig. 38-28. (*a*) Calculate E, B for $a < r < b$. (*b*) Calculate the Poynting vector $\vec{\mathbf{S}}$ for $a < r < b$. (*c*) By suitably integrating the Poynting vector, show that the total power flowing across the annular cross section $a < r < b$ is $\mathscr{E}^2/R$. Is this reasonable? (*d*) Show that the direction of $\vec{\mathbf{S}}$ is always from the battery to the resistor, no matter which way the battery is connected.

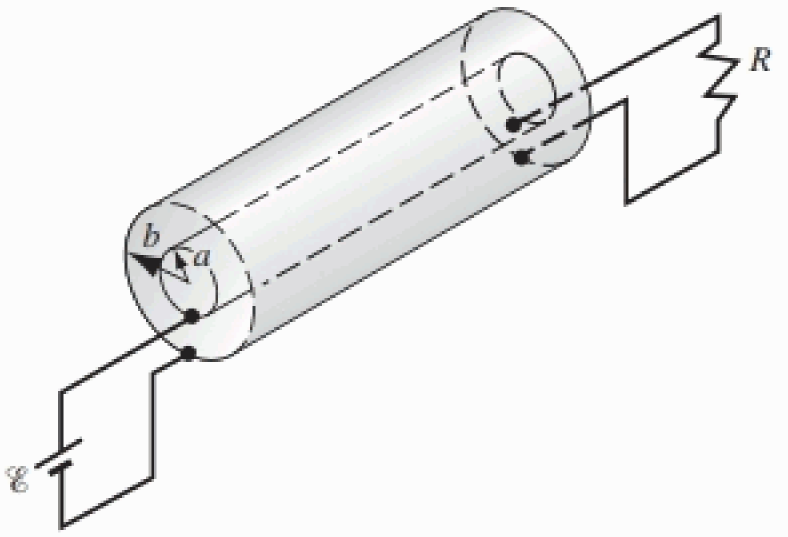


FIGURE 38-28. Problem 11.

12. Figure 38-29 shows a cylindrical resistor of length l, radius a, and resistivity ρ, carrying a current i. (*a*) Show that the Poynting vector $\vec{\mathbf{S}}$ at the surface of the resistor is everywhere directed normal to the surface, as shown. (*b*) Show that the rate at which energy flows into the resistor through its cylindrical surface, calculated by integrating the Poynting vector over this surface, is equal to the rate at which internal energy is produced; that is,

$$\int \vec{\mathbf{S}} \cdot d\vec{\mathbf{A}} = i^2 R,$$

where $d\vec{\mathbf{A}}$ is an element of area of the cylindrical surface. This suggests that, according to the Poynting vector point of view, the energy that appears in a resistor as internal energy does not enter it through the connecting wires but through the space around the wires and the resistor.

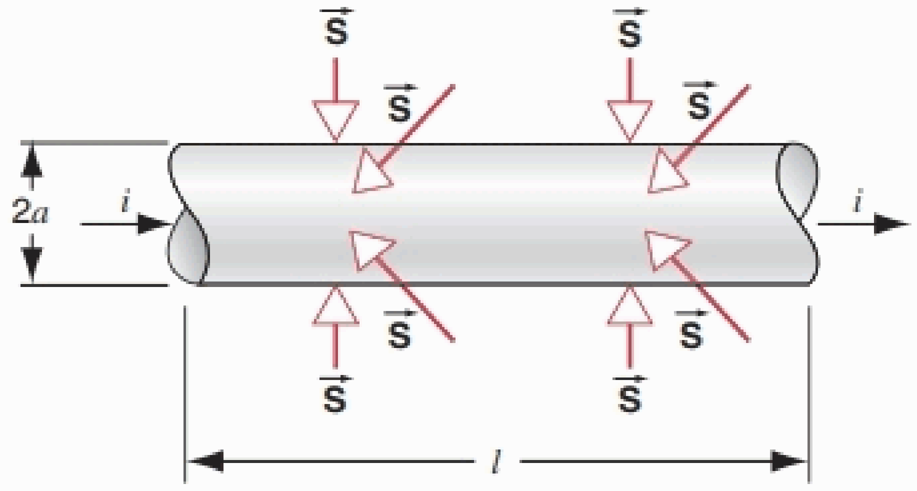


FIGURE 38-29. Problem 12.

13. A plane electromagnetic wave, with wavelength 3.18 m, travels in free space in the $+x$ direction with its electric vector $\vec{\mathbf{E}}$, of amplitude 288 V/m, directed along the y axis. (*a*) What is the frequency of the wave? (*b*) What is the direction and amplitude of the magnetic field associated with the wave? (*c*) If $E = E_m \sin(kx - \omega t)$, what are the values of k and ω? (*d*) Find the intensity of the wave. (*e*) If the wave falls on a perfectly absorbing sheet of area 1.85 m^2, at what rate would momentum be delivered to the sheet and what is the radiation pressure exerted on the sheet?

14. Figure 38-30 shows a parallel-plate capacitor being charged. (*a*) Show that the Poynting vector $\vec{\mathbf{S}}$ points everywhere radially into the cylindrical volume. (*b*) Show that the rate at which energy flows into this volume, calculated by integrating the Poynting vector over the cylindrical boundary of this volume, is equal to the rate at which the stored electrostatic energy increases; that is,

$$\int \vec{\mathbf{S}} \cdot d\vec{\mathbf{A}} = Ad \frac{d}{dt}\left(\frac{1}{2}\epsilon_0 E^2\right),$$

where Ad is the volume of the capacitor and $\frac{1}{2}\epsilon_0 E^2$ is the energy density for all points within that volume. This analysis shows that, according to the Poynting vector point of view, the energy stored in a capacitor does not enter it through the wires but through the space around the wires and the plates. (Hint: To find $\vec{\mathbf{S}}$ we must first find $\vec{\mathbf{B}}$, which is the magnetic field set up by the displacement current during the charging process; see Fig. 38-2. Ignore fringing of the lines of $\vec{\mathbf{E}}$.)

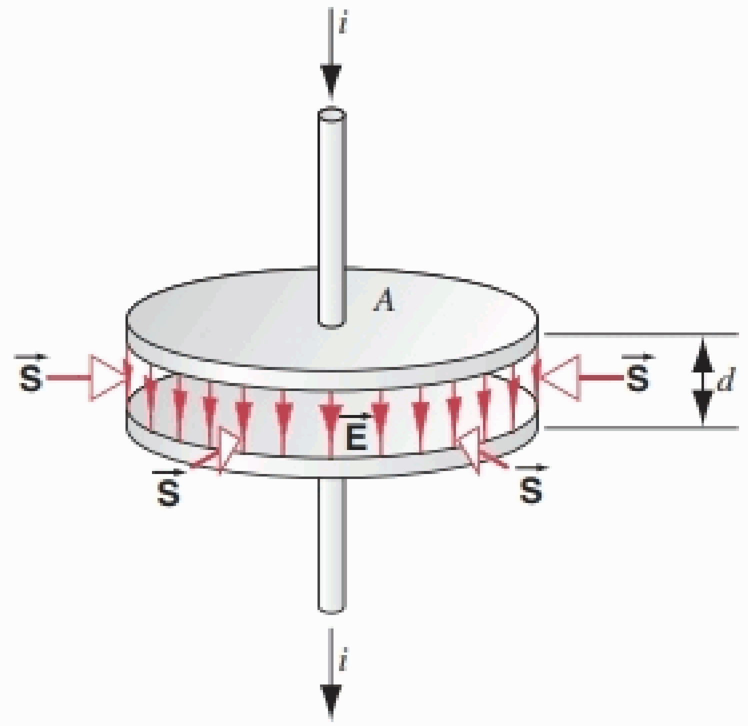


FIGURE 38-30. Problem 14.

15. A helium–neon laser of the type often found in physics laboratories has a beam power output of 5.00 mW at a wavelength of 633 nm. The beam is focused by a lens to a circular spot whose effective diameter may be taken to be 2.10 wavelengths. Calculate (*a*) the intensity of the focused beam, (*b*) the radiation pressure exerted on a tiny, perfectly absorbing sphere whose diameter is that of the focal spot, (*c*) the force exerted on this sphere, and (*d*) the acceleration imparted to it. Assume a sphere density of 4.88 g/cm^3.

16. It has been proposed that a spaceship might be propelled in the solar system by radiation pressure, using a large sail made of foil. How large must the sail be if the radiation force is to be equal in magnitude to the Sun's gravitational attraction? Assume that the mass of the ship + sail is 1650 kg, that the sail is perfectly reflecting, and that the sail is oriented at right angles to the Sun's rays. See Appendix C for needed data.

CHAPTER 39

LIGHT WAVES

In this chapter we discuss the characteristics of light waves, including sources of visible radiation, the speed of propagation in vacuum and in matter, the change in direction that occurs when light encounters a boundary between two materials in which the speed of propagation is different, and the Doppler effect (the change in frequency due to relative motion of source and observer).

Because there is nothing fundamental that distinguishes light from any other type of electromagnetic wave, you should keep in mind that the discussions in this chapter could apply equally well to other kinds of electromagnetic waves. Nevertheless, this chapter serves as a bridge from our discussion of electromagnetic waves in the previous chapter to the discussion of optics (the science of light), which will be the subject of the next several chapters.

39-1 THE ELECTROMAGNETIC SPECTRUM*

In Maxwell's time, light and the adjoining infrared and ultraviolet radiation were the only known types of electromagnetic radiation. Today the electromagnetic spectrum, shown in Fig. 39-1, includes a broad range of different kinds of radiation from a variety of sources. From Maxwell's theory we conclude that, even though these radiations differ greatly in their properties, in their means of production, and in the ways we observe them, they share other features in common: they all can be described in terms of electric and magnetic fields, and they all travel through vacuum with the same speed (the speed of light). In fact, from the fundamental point of view, they differ *only* in wavelength or frequency. The names given to the various regions of the spectrum in Fig. 39-1 have to do only with the way the different types of waves are produced or observed; they have nothing to do with any fundamental property of the waves. Other than the difference in their wavelengths, there is no experimental way to distinguish a wave in the visible region from one in the infrared region; the waves have identical forms and identical mathematical descriptions. There are no gaps in the spectrum, nor are there sharp boundaries between the various categories. (Certain regions of the spectrum are assigned by law for commercial or other uses, such as TV, AM, or FM broadcasting.)

Let us consider some of these types of electromagnetic radiation in more detail.

1. *Light.* The visible region of the spectrum is the one most familiar to us, because as a species we have adapted receptors (eyes) that are sensitive to the most intense electromagnetic radiation emitted by the Sun, the closest extraterrestrial source. The limits of the wavelength of the

---

*The word *spectrum* comes from a Latin word meaning "form" or "appearance." Other familiar words from the same root include "spectacle" and "species." Newton introduced the word to describe the rainbow-like image that resulted when a beam of sunlight passed through a glass prism. Today we speak of the *electromagnetic spectrum* to indicate the many different kinds of electromagnetic radiation, classified according to their frequency or wavelength on a scale from small to large. We also speak of the *political spectrum,* which similarly indicates the broad range of political views on a scale from ultraconservative to ultraliberal.

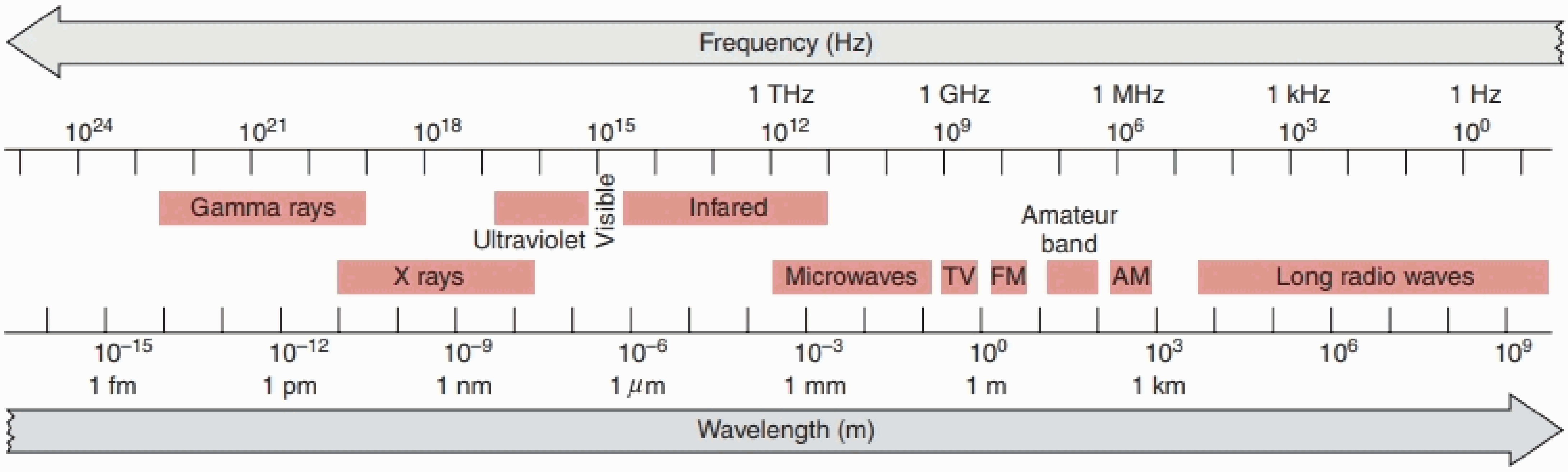


FIGURE 39-1. The electromagnetic spectrum. Note that both the wavelength and frequency scales are logarithmic.

visible region are from about 400 nm (violet) to about 700 nm (red).

Light can be emitted from a dilute collection of atoms, such as a gas, in which the light is characteristic of the properties of individual atoms, or from a concentrated collection of atoms, such as a solid, in which the light may be characteristic of the properties of the entire object. The study of the light emitted by the Sun and by distant stars can give us information about their chemical composition, temperature, and motion.

2. *Infrared rays.* Infrared radiation, which has wavelengths longer than the visible (from 0.7 μm to about 1 mm), is commonly emitted by atoms or molecules when they change their rotational or vibrational motion. Infrared radiation is sometimes called *heat radiation,* because it is associated with heat transfer by radiation as an object gains or loses internal energy; objects with temperatures in the range we normally encounter (up to about 3000 K) emit their most intense radiation in the infrared region. Mapping the infrared radiation from space has given us information that supplements that obtained from visible radiation (Fig. 39-2).

3. *Microwaves.* Microwaves can be regarded as short radio waves, with typical wavelengths in the range 1 mm to 1 m. They are commonly produced by oscillating electric circuits, as in the case of microwave ovens. Microwaves are commonly used to relay telephone calls (Fig. 39-3). On the cosmic scale, the universe is filled with radiation that is the remnant of the "big bang" event that marked the origin of the universe some 10^{10} years ago (see Section 52-5); this radiation is mostly in the microwave region of the spectrum.

4. *Radio waves.* Radio waves have wavelengths longer than 1 m. They can be produced by electrons oscillating in wires of electric circuits, and antennas can be used to transmit or receive radio waves that carry AM or FM radio and TV signals. There are also extraterrestrial sources of radio waves; the most intense is the Sun, whose radio wave emissions can interfere with radio or TV reception on Earth. The planet Jupiter is also an active source of radio emissions. Observing such emissions, which is called *radio astronomy,* provides us with additional information about the universe. Figure 39-4 shows an example of a radiotelescope, and Fig. 39-5 shows a map of radio emissions of our galaxy.

5. *Ultraviolet rays.* Radiations with wavelengths shorter than the visible begin with the ultraviolet (1 nm to 400 nm), which can be produced by electrons in atoms as well as by thermal sources such as the Sun. Exposure to ultraviolet radiation can cause sunburn and even skin cancers,

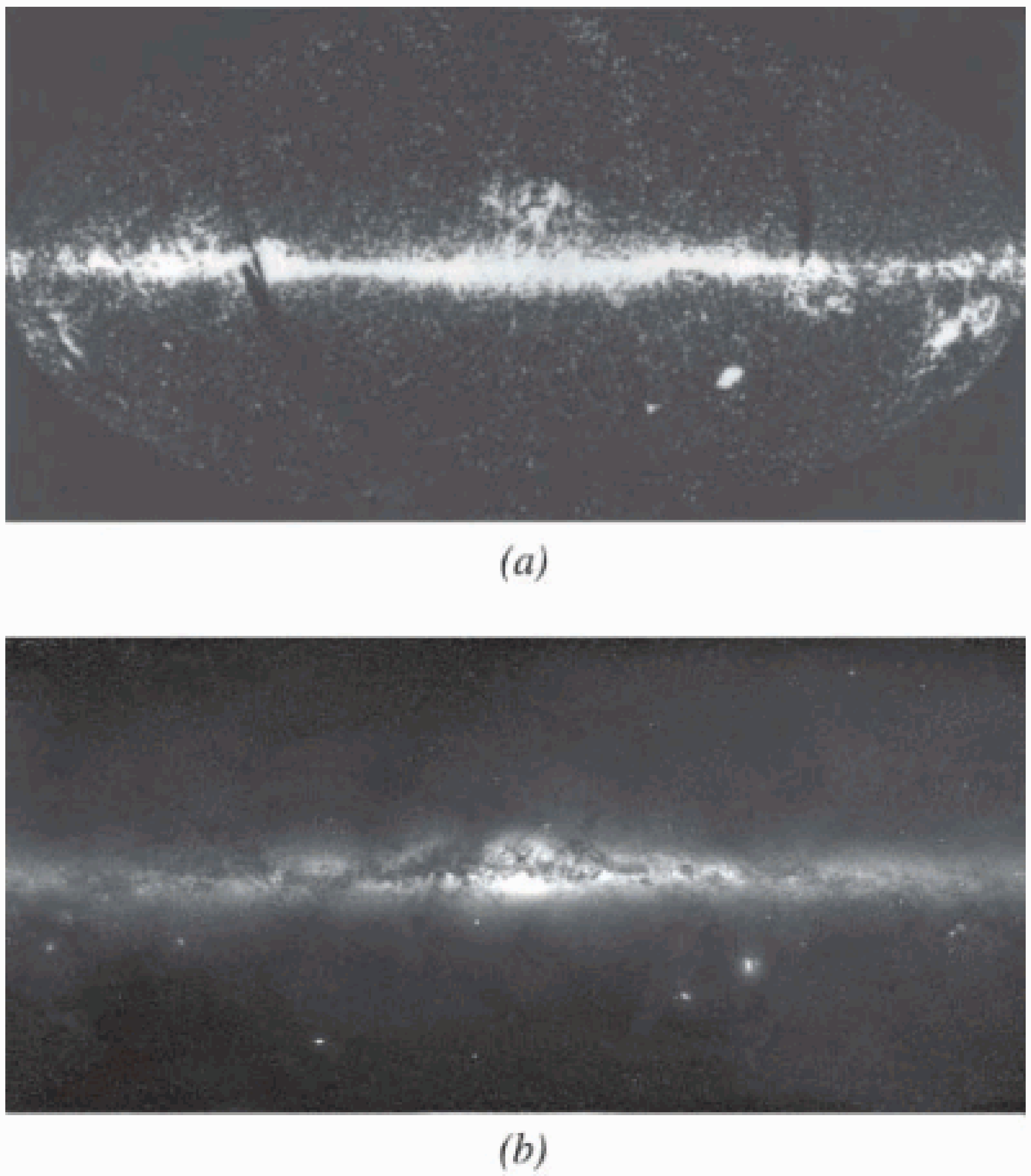


FIGURE 39-2. (*a*) Infrared image of our Milky Way galaxy taken by the IRAS satellite. (*b*) Visible-light image of the Milky Way. Parts of the visible image, especially those near the center of the galaxy, are obscured by dust clouds, which do not affect the infrared image. The two large objects below the galaxy and right of center are the Large and Small Magellanic Clouds, which are companion galaxies to the Milky Way.

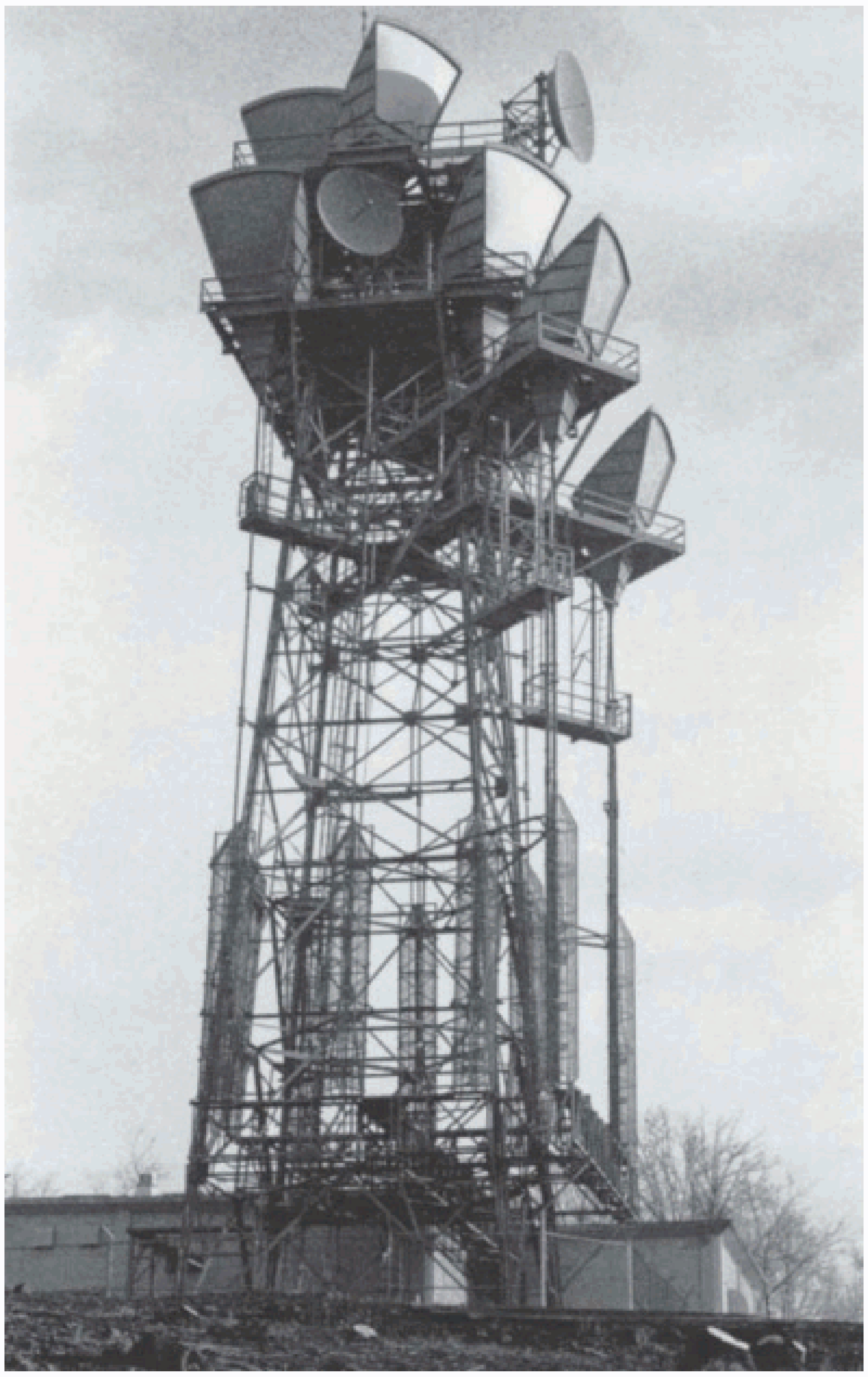


FIGURE 39-3. A microwave relay station, which receives and then re-transmits signals that carry long-distance telephone calls.

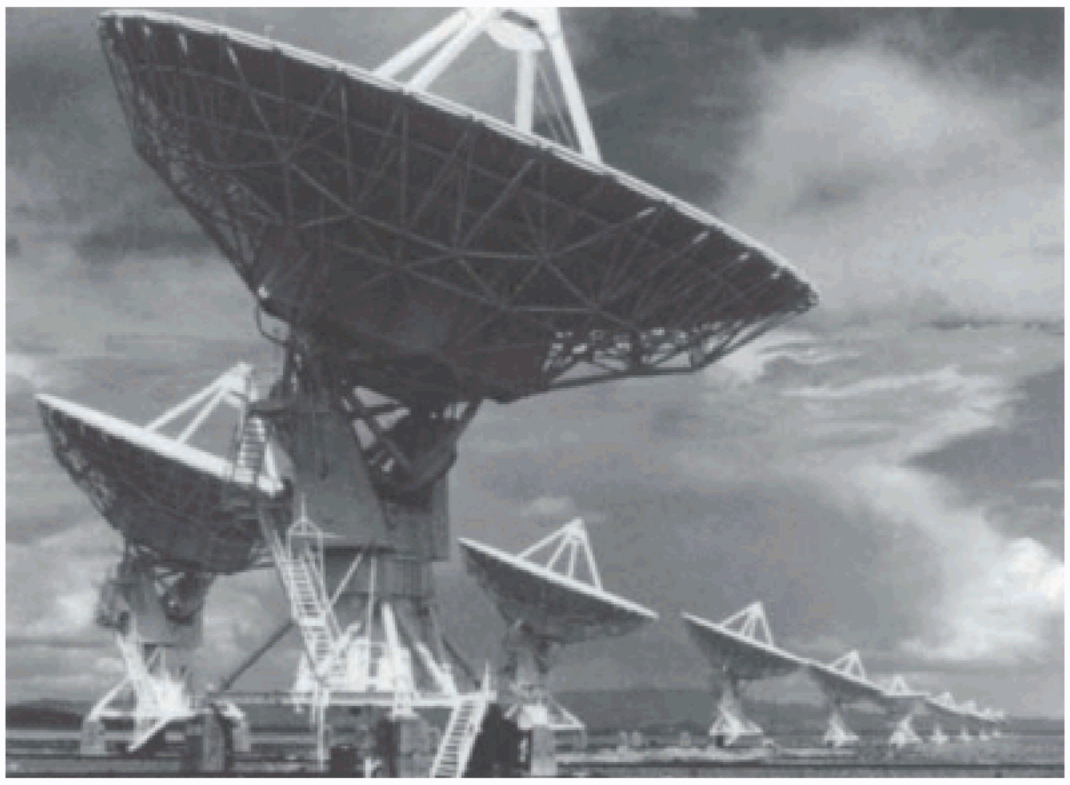


FIGURE 39-4. One of the 27 25-m-diameter radiotelescope antenna dishes at the Very Large Array near Socorro, New Mexico. The 27 dishes are arranged on a Y-shaped railroad track, each leg of which is 10 miles long. This arrangement is equivalent to a single dish 20 miles in diameter.

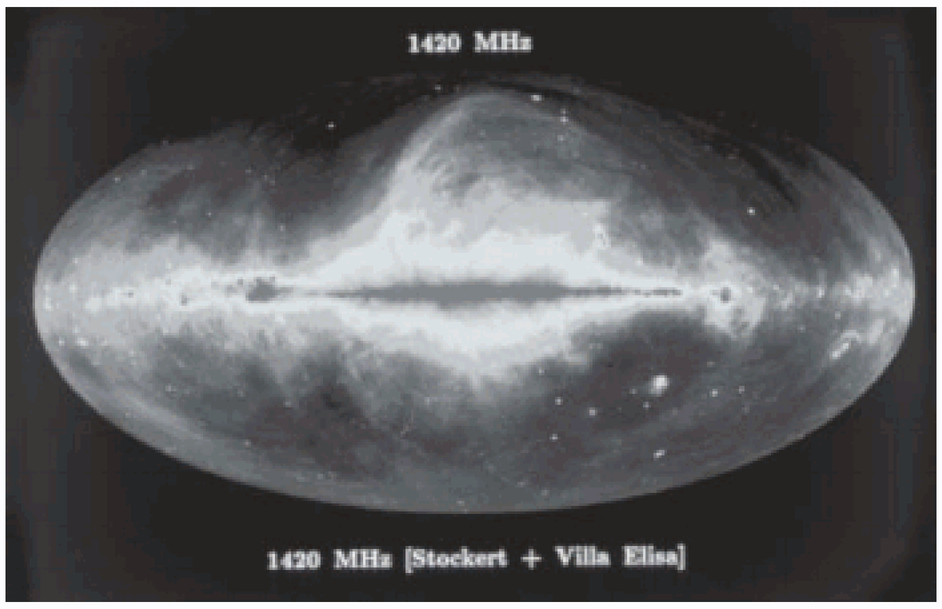


FIGURE 39-5. A radio image of the Milky Way galaxy. (Compare with Fig. 39-2.) This image was taken at a wavelength of 73 cm. This radiation mostly originates from high-energy electrons that are deflected by magnetic fields in the galaxy. Note the intense emissions out of the plane of the galaxy, which do not appear in Fig. 39-2.

so it is fortunate that most of the ultraviolet radiation from the Sun is absorbed by Earth's atmosphere. Because of this absorption, ultraviolet astronomy must rely on data from observatories in Earth orbit.

6. *X rays.* X rays (typical wavelengths from 0.01 nm to 10 nm) can be produced with discrete wavelengths in atoms by transitions involving the most tightly bound electrons, and they can also be produced in a continuous range of wavelengths when charged particles such as electrons are decelerated. Because x rays can penetrate soft tissue but not bone, they find wide use in medical diagnosis. Ordinarily, stars are not strong x-ray emitters, but in binary star systems material from one star can emit x rays when it is heated and accelerated as it falls into the other star.

7. *Gamma rays.* Gamma rays, which have the shortest wavelengths in the electromagnetic spectrum (less than 10 pm), are emitted in the decays of many radioactive nuclei and certain elementary particles. They are the most penetrating form of electromagnetic radiation, and exposure to intense gamma radiation can have a harmful effect on the human body. In much smaller doses, radioactive materials can be introduced into the body to allow physicians to form an image of a particular organ by observing the emitted gamma radiations. Each type of radioactive nucleus emits its own unique pattern of gamma radiation, and these emissions can be used to identify the presence of specific elements. In gamma-ray astronomy, observation of these characteristic gamma radiations from stars and galaxies serves as evidence for the presence of particular elements and for the existence of particular nuclear processes in the universe.

From the above descriptions, you can see that there are both natural and artificial sources of all types of electromagnetic radiation, and you can also see that the study of electromagnetic radiations at all wavelengths can be used to add to our knowledge about the structure and evolution of the universe.

Our study of light in the next several chapters will concentrate on its wave nature. However, some experiments suggest that light can behave not as a smoothly varying wave but rather as a concentrated bundle of electromagnetic energy known as a *photon.* We will consider the photon aspect of light beginning in Chapter 45.

39-2 VISIBLE LIGHT

We may operationally define visible light to be electromagnetic radiation to which the eye is sensitive. The sensitivities of individual observers may vary, but humans can typically observe radiation with wavelengths from 400 nm to 700 nm (corresponding to frequencies from 7×10^{14} Hz to 4×10^{14} Hz). Within that range, the sensitivity to different wavelengths is not at all constant. Figure 39-6 shows a representation of the variation in the sensitivity of an average human observer to radiations of differing wavelength but constant radiant intensity over the visible region of the spectrum. The greatest sensitivity occurs near 555 nm, corresponding to light of a yellow-green color. The sensitivity decreases rapidly at both long and short wavelengths; we can roughly set the limits of the visible region, where the sensitivity is equal to 1% of that of the peak, to be 430 nm (violet) and 690 nm (red).* (Compare the range of *visible* wavelengths, less than a factor of 2, with the range of *audible* wavelengths or frequencies, which Fig. 19-5 shows to be about a factor of 100 at the 1% limit.)

Sources of visible light depend ultimately on the motion of electrons. Electrons in atoms can be raised from their lowest energy state to higher states by various means, such as by heating the substance or by passing an electric current through it. When the electrons eventually drop back to their lowest levels, the atoms emit radiation that may be in the visible region of the spectrum. Emission of visible light is particularly likely when the outer (valence) electrons are the ones making the transitions.

The most familiar source of visible light is the Sun. Its surface emits radiation across the entire electromagnetic spectrum, but its most intense radiation is in the region we define as visible, and the Sun's radiant intensity peaks at a wavelength of about 550 nm, corresponding precisely to the peak in the sensitivity of the average human observer (Fig. 39-6). This suggests that, through natural selection, our eyes evolved in such a way that their sensitivity matched the Sun's spectrum.

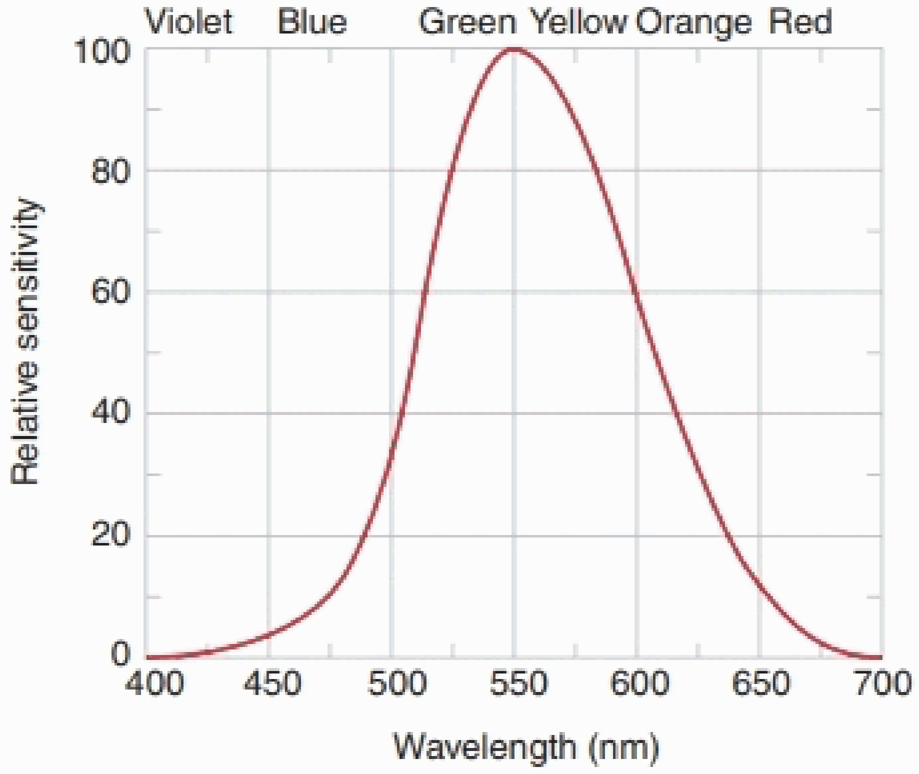


FIGURE 39-6. The relative sensitivity of the human eye as a function of wavelength.

All objects emit electromagnetic radiation, called *thermal radiation,* because of their temperature. Objects such as the Sun, whose thermal radiation is visible, are called *incandescent.* Other common incandescent objects are the filaments of ordinary lightbulbs and the glowing embers in a charcoal fire. Incandescence is normally associated with hot objects; typically, temperatures in excess of 1000°C are required.

It is also possible for light to be emitted from cool objects; this phenomenon is called *luminescence.* Examples include common fluorescent lamps, lightning, glowing watch and clock dials, and television receivers. In the case of a fluorescent lamp, an electric current passed through the gas in the tube causes the electrons to move to higher energy states; when the electrons return to their original energy states, they give up their excess energy in the form of ultraviolet radiation. This radiation is absorbed by atoms of the coating on the inside of the glass tube, which then emit visible light. In the case of glowing clock dials, it is *incident* light that causes the excitation.

Luminescent objects can be put into two categories depending on the duration of light emission after the source of excitation is removed. Objects in which the emission of light ceases immediately (within 10^{-8} s) after the excitation is removed are called *fluorescent,* for example, the fluorescent lamp. Objects that continue to glow longer than 10^{-8} s after the source of the excitation is removed (such as the clock dial) are called *phosphorescent,* and the material that causes this effect is called a *phosphor* (Fig. 39-7).

Luminescence can have a variety of causes. When the energy that excites the atoms originates from a chemical reaction, it is called *chemiluminescence.* Often the effect occurs in living things, such as in fireflies and many marine organisms, in which case it is called *bioluminescence* (Fig. 39-8). Light can also be emitted when certain crystals—for example, sugar—are crushed; the effect, called *triboluminescence,* can be observed in a dark room by crunching

---

*The assignment of color to the various regions of the visible spectrum is quite arbitrary, because color is primarily a psychological label rather than a physical quality. Just as there is no fundamental physical distinction between light and other electromagnetic waves, there is no fundamental physical distinction between blue light and red light. For more on the perception of color, see "The Retinex Theory of Color Vision," by Edwin H. Land, *Scientific American,* December 1977, p. 108, and *Eye, Brain, and Vision,* by David H. Hubel (Scientific American Library Series, 1988), Chapter 8.

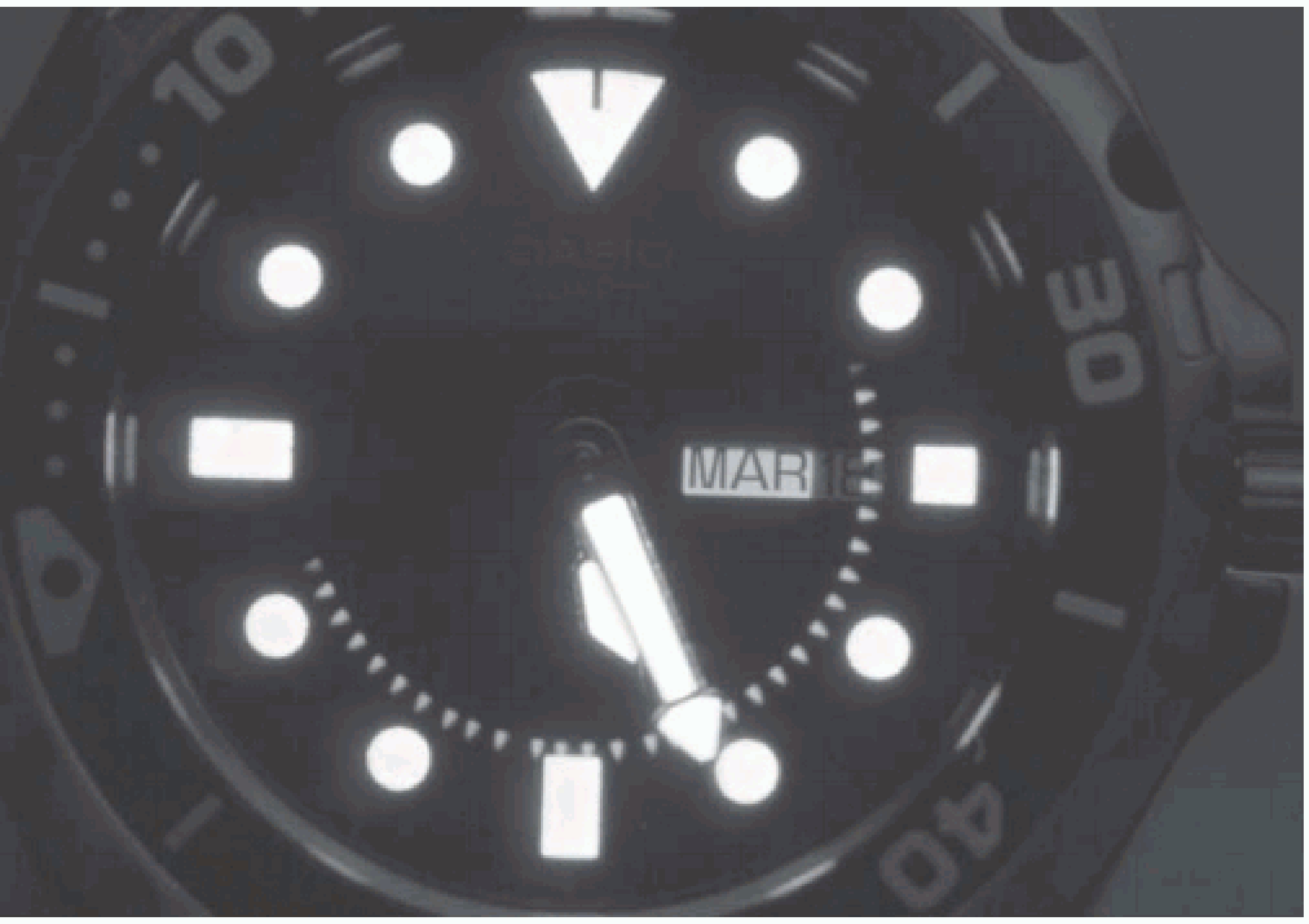


FIGURE 39-7. A phosphorescent material causes the numbers on this watch to glow. Radioactive radium is often mixed with the phosphor to provide the excitation energy.

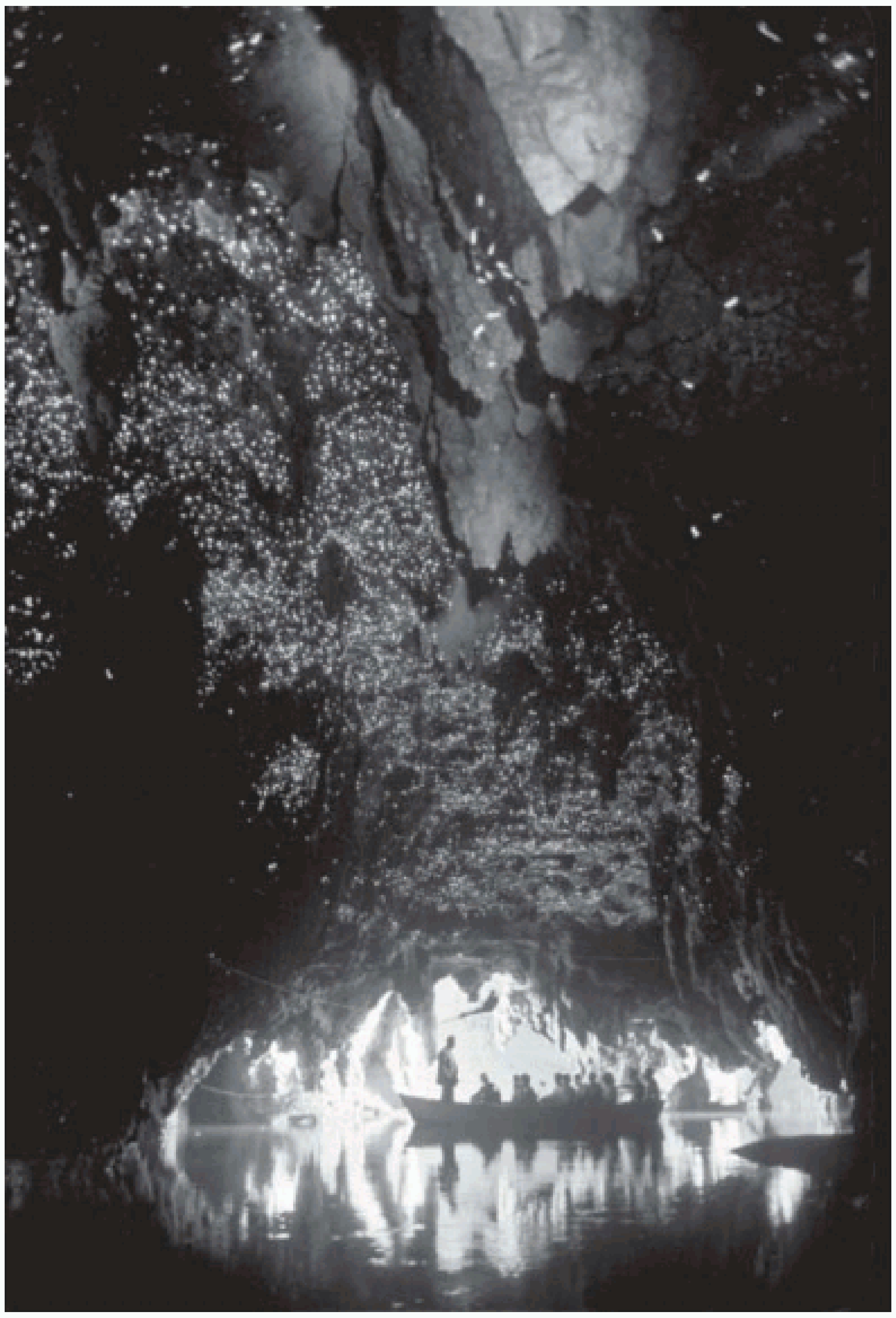


FIGURE 39-8. The dots of light are glow-worms in a cave in New Zealand. The light attracts insects, which are trapped and serve as food for the glow-worm larvae.

Wintergreen Life-Savers™ between the teeth. Other causes of luminescence include electric currents (as in lightning or light-emitting diodes) and the impact of high-energy particles (as in the aurora borealis).

39-3 THE SPEED OF LIGHT

According to Maxwell's theory, all electromagnetic waves travel through empty space with the same speed. We call this speed "the speed of light," even though it applies to all electromagnetic radiations, not only light. This speed is one of the fundamental constants of nature.

Until the 17th century, it was generally believed that light propagated instantaneously; that is, the speed of light was thought to be infinite. Galileo discussed this question in his famous work, *Dialogue Concerning Two New Sciences,* published in 1638. He presented his arguments in the form of a dialogue between several characters, including Simplicio (representing the scientifically ignorant) and Sagredo (representing the voice of reason and probably Galileo himself):

> SIMPLICIO: Everyday experience shows that the propagation of light is instantaneous; for when we see a piece of artillery fired, at a great distance, the flash reaches our eyes without a lapse of time, but the sound reaches the ear only after a noticeable interval.
>
> SAGREDO: Well, Simplicio, the only thing I am able to infer from this familiar bit of experience is that sound, in reaching our ear, travels more slowly than light; it does not inform me whether the coming of the light is instantaneous or whether, although extremely rapid, it still occupies time.

Galileo then goes on to describe an experiment (which he actually carried out) to measure the speed of light. He and an assistant stood facing one another at night, separated by a distance of about a mile, each carrying a lantern that could be covered or uncovered at will. Galileo started by uncovering his lantern, and the assistant was to uncover his lantern when he saw the light from Galileo's. Galileo then tried to measure the time interval between the instant at which he uncovered his lantern and the instant at which the light from his assistant's lantern reached him. Although Galileo was not able to determine a value for the speed of light (the round-trip time for a separation of 1 mile being only 11 μs, several orders of magnitude smaller than human reaction times), he is credited with the first attempt at measuring the speed of light.

In 1676, Ole Roemer, a Danish astronomer working in Paris, used astronomical observations to deduce that the speed of light is finite. His conclusion was based on a discrepancy between the predicted and observed times of eclipses of Jupiter's innermost moon, Io (see Problem 2). About 50 years later James Bradley, an English astronomer,

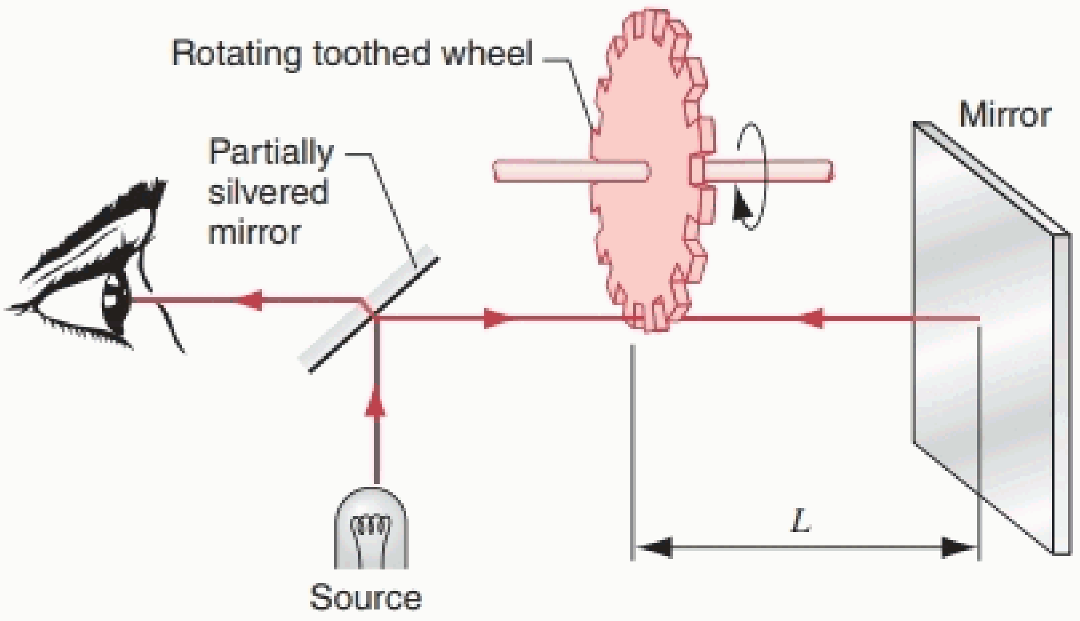


FIGURE 39-9. A schematic diagram of Fizeau's apparatus for measuring the speed of light.

used a different technique based on starlight to obtain a value of 3×10^8 m/s.

The next major improvement in measuring the speed of light did not come for more than a century. In 1849, the French physicist Hippolyte Louis Fizeau (1819–1896) used a mechanical arrangement, illustrated in Fig. 39-9. In essence, a light beam was made to travel a long round-trip path (of length $L = 8630$ m each way), passing through a rotating toothed wheel in each direction. The rotating wheel chops the beam going toward the mirror into short pulses. If, during the time the pulse travels the round trip to the mirror and back, the wheel rotates so that a tooth is now blocking the light path, the observer does not see the light pulse. When this occurs, the time $2L/c$ it takes the light beam to make the round trip between the wheel and the mirror must equal the time θ/ω it takes for the wheel to rotate at angular speed ω through the angle θ between the center of a tooth and the center of a gap. That is,

$$\frac{2L}{c} = \frac{\theta}{\omega}$$

or

$$c = \frac{2L\omega}{\theta}. \tag{39-1}$$

Chopped beams are used in similar ways to measure the speeds of neutrons and other particles. (A variant of this method was used to verify the Maxwell speed distribution; see Fig. 22-8.)

Fizeau's result using this method was 3.133×10^8 m/s. Other experimenters, including the U.S. physicist Albert A. Michelson, used similar mechanical techniques throughout the late 19th and early 20th centuries. Michelson's work was noteworthy for its care and precision, and he was awarded the 1907 Nobel Prize in physics for his research using optical techniques to make precise measurements. As a result of these investigations, the uncertainty in c was reduced to about 1000 m/s.

The development of electronic techniques, especially as applied to microwaves, permitted a new class of measurements to be done in the 1950s. These measurements gave results that agreed with Michelson's and had similar limits of uncertainty.

The breakthrough in measurements of the speed of light came in the 1970s with the application of lasers. By measuring the frequency and wavelength directly, the speed of light could be obtained from $c = \lambda f$. Refinements of this technique have resulted in values of c with uncertainties smaller than 1 m/s. Table 39-1 summarizes some of the

TABLE 39-1 **Speed of Electromagnetic Radiation in Free Space (some selected measurements)**

Date	*Experimenter*	*Country*	*Method*	*Speed* (km/s)	*Uncertainty* (km/s)
1600 (?)	Galileo	Italy	Lanterns and shutters	"Extraordinarily rapid"	
1676	Roemer	France	Moons of Jupiter	"Finite"	
1729	Bradley	England	Aberration of starlight	304,000	
1849	Fizeau	France	Toothed wheel	313,300	
1862	Foucault	France	Rotating mirror	298,000	500
1880	Michelson	United States	Rotating mirror	299,910	50
1906	Rosa and Dorsey	United States	Electromagnetic theory	299,781	10
1923	Mercier	France	Standing waves on wires	299,782	15
1926	Michelson	United States	Rotating mirror	299,796	4
1950	Bergstrand	Sweden	Geodimeter	299,792.7	0.25
1950	Essen	England	Microwave cavity	299,792.5	3
1950	Bol and Hansen	United States	Microwave cavity	299,789.3	0.4
1951	Aslakson	United States	Shoran radar	299,794.2	1.9
1952	Rank et al.	United States	Molecular spectra	299,766	7
1952	Froome	England	Microwave interferometer	299,792.6	0.7
1958	Froome	England	Microwave interferometer	299,792.50	0.10
1967	Grosse	Germany	Geodimeter	299,792.5	0.05
1973	Evenson et al.	United States	Laser techniques	299,792.4574	0.0012
1978	Woods et al.	England	Laser techniques	299,792.4588	0.0002
1987	Jennings et al.	United States	Laser techniques	299,792.4586	0.0003

measurements of c we have discussed.* Note the reduction in the limit of uncertainty over the years.

The precision of measuring frequency (about 1 part in 10^{13}) has far exceeded that of measuring wavelength (about 1 part in 10^9). As a result, we now define the speed of light to have the exact value

$$c = 299{,}792{,}458 \text{ m/s},$$

and the second is defined based on measurements of frequency, so that the meter is now a secondary standard, defined in terms of the second and the value of c.

The Speed of Light in Matter

When we refer to "the speed of light," we usually mean the speed *in vacuum.* We have discussed in Chapter 38 the propagation of electromagnetic radiation, which takes place through the coupling between its electric and magnetic fields. In dielectric materials, we have seen in Section 29-6 that the electric field is altered by a factor of κ_e, the dielectric constant of the material. A convenient way of modifying equations for electric fields in vacuum to account for the presence of dielectric materials is, as shown in Section 29-6, to replace the electric constant ϵ_0 with the quantity $\kappa_e\epsilon_0$.

We also must account for the effect of the *magnetic* properties of the medium on the magnetic field of the propagating electromagnetic wave. As we discussed in Section 35-4, magnetic materials are characterized by a relative permeability κ_m, and in analogy with the electric field we can modify the magnetic field equations in matter by replacing the magnetic constant μ_0 with the quantity $\kappa_m\mu_0$.

Making these substitutions, we can therefore modify Eq. 38-19 to give the speed of light in matter:

$$v = \frac{1}{\sqrt{\kappa_m\kappa_e}}\frac{1}{\sqrt{\mu_0\epsilon_0}} = \frac{c}{\sqrt{\kappa_m\kappa_e}}. \qquad (39\text{-}2)$$

Materials that transmit light are normally nonferromagnetic, and therefore κ_m differs from 1 typically by no more than 10^{-4} (see Tables 35-2 and 35-3). It is therefore the dielectric constant κ_e that determines the speed of light in a material. However, the dielectric constants that are listed in Table 29-2 cannot be used in Eq. 39-2, because those values are characteristic of *static* situations. Recall that the dielectric constant is in effect a measure of the response of the dipoles (permanent or induced) to an applied electric field. If the applied field varies at high frequency, the dipoles may not have time to respond, and we cannot use the static dielectric constants in the case of a rapidly varying $\vec{\mathbf{E}}$ field. At the frequencies characteristic of a light wave (10^{15} Hz), the field oscillates too rapidly for the dipoles to follow completely. Furthermore, κ_e in Eq. 39-2 varies with frequency, so that the speed of light in matter depends on the wavelength or frequency of the light.

Table 39-2 shows values of the speed of light in various materials.

TABLE 39-2 Speed of Light in Selected Materials[a]

Material	*Speed of Light* (10^8 m/s)
Vacuum	3.00
Air	3.00
Water	2.26
Sugar solution (50%)	2.11
Crown glass	1.97
Diamond	1.24

[a] For a wavelength of 589 nm (yellow sodium light).

Propagation of Light in Matter (Optional)

The mechanism responsible for the propagation of light in matter is scattering (in effect, absorption of the incident light by the atoms or molecules of the medium and reemission of the light in various directions). The phases of these scattered waves traveling transverse to the direction of the incident light cause nearly complete destructive interference in the transverse directions. The scattered waves traveling parallel to the direction of the incident light are not in phase with the incident light; as a result of the interference between the two waves, the phase of their combination differs from the phase of the incident wave. We observe this change in phase as a change in speed.

The electric field of the incident light causes the electrons in an atom to oscillate with the frequency of the incident light. It is reasonable to expect that the phase of the reemitted wave depends on the frequency of the atomic oscillation and therefore on the frequency of the original wave. When the incident and scattered waves interfere, the phase of their combination depends on their phase difference and hence on the frequency. As a result, the speed of light in a material depends on the frequency or wavelength. This phenomenon, which is called *dispersion,* is discussed in Section 39-4.

In a typical solid, the distance over which the original light is absorbed and reemitted is of the order of micrometers, and in air it is of the order of millimeters. In effect, the light that we see from the Sun comes to our eyes not directly from the Sun but from the molecules of air a few millimeters in front of our eyes. ■

SAMPLE PROBLEM 39-1. The speed of light of yellow color ($\lambda = 589$ nm, corresponding to $f = 5.09 \times 10^{14}$ Hz) in water is 2.26×10^8 m/s. What is the effective dielectric constant for water at this frequency?

Solution We use Eq. 39-2 and assume that, to sufficient accuracy for this calculation, $\kappa_m = 1$. Solving Eq. 39-2 for κ_e and putting $\kappa_m = 1$, we obtain

$$\kappa_e = \left(\frac{c}{v}\right)^2 = \left(\frac{3.00 \times 10^8 \text{ m/s}}{2.26 \times 10^8 \text{ m/s}}\right)^2 = 1.76.$$

---

*For references to some of these measurements, see "Resource Letter RMSL-1: Recent Measurements of the Speed of Light and The Redefinition of the Meter," by Harry E. Bates, *American Journal of Physics,* August 1988, p. 682.

This is very different from the *static* dielectric constant for water, which has a value of about 80 at room temperature, suggesting the difficulty that the dipole moments of water molecules have in following the variation of the electric field at this frequency. In general, dielectric constants at high frequency are smaller than the corresponding static values, which means that the high-frequency induced electric field is smaller than the static induced electric field.

39-4 REFLECTION AND REFRACTION OF LIGHT WAVES

When you look at a pane of window glass, you of course notice that light reaches you from the other side of the glass, and a friend standing on the other side of the glass is able to see you. If you look carefully, however, you may also see your reflection in the glass. If you were to shine a flashlight on the glass, your friend would see the beam of light, but you might also see some of the light reflected back toward you.

In general, these two effects can occur whenever a beam of light travels from one medium (the air, for instance) to another (the glass). Part of the beam may be reflected back into the first medium, and part may be transmitted into the second medium. Figure 39-10 illustrates these two effects. Note that the beam of light may be bent or *refracted* as it enters the second medium.*

In Fig. 39-10, the beams are represented by rays. The rays, which are drawn as straight lines perpendicular to the (plane) wavefronts, indicate the direction of motion of the wavefronts. Note the three rays shown in Fig. 39-10: the original or *incident* ray, the *reflected* ray, and the *refracted* ray, which changes direction as it enters the second medium.

At the point where the incident ray strikes the surface, we draw a line normal (perpendicular) to the surface, and we define three angles measured with respect to the normal: the *angle of incidence* θ_1, the *angle of reflection* θ_1', and the *angle of refraction* θ_2. (The subscripts on the angles indicate the medium through which the ray travels. In our case, the ray is incident from medium 1, the air, and enters medium 2, the glass.) The plane formed by the incident ray and the normal is called the *plane of incidence;* it is the plane of the page in Fig. 39-10.

From experiment, we deduce the following laws governing reflection and refraction:

The Law of Reflection The reflected ray lies in the plane of incidence, and

$$\theta_1' = \theta_1. \tag{39-3}$$

The Law of Refraction The refracted ray lies in the plane of incidence, and

$$n_1 \sin \theta_1 = n_2 \sin \theta_2. \tag{39-4}$$

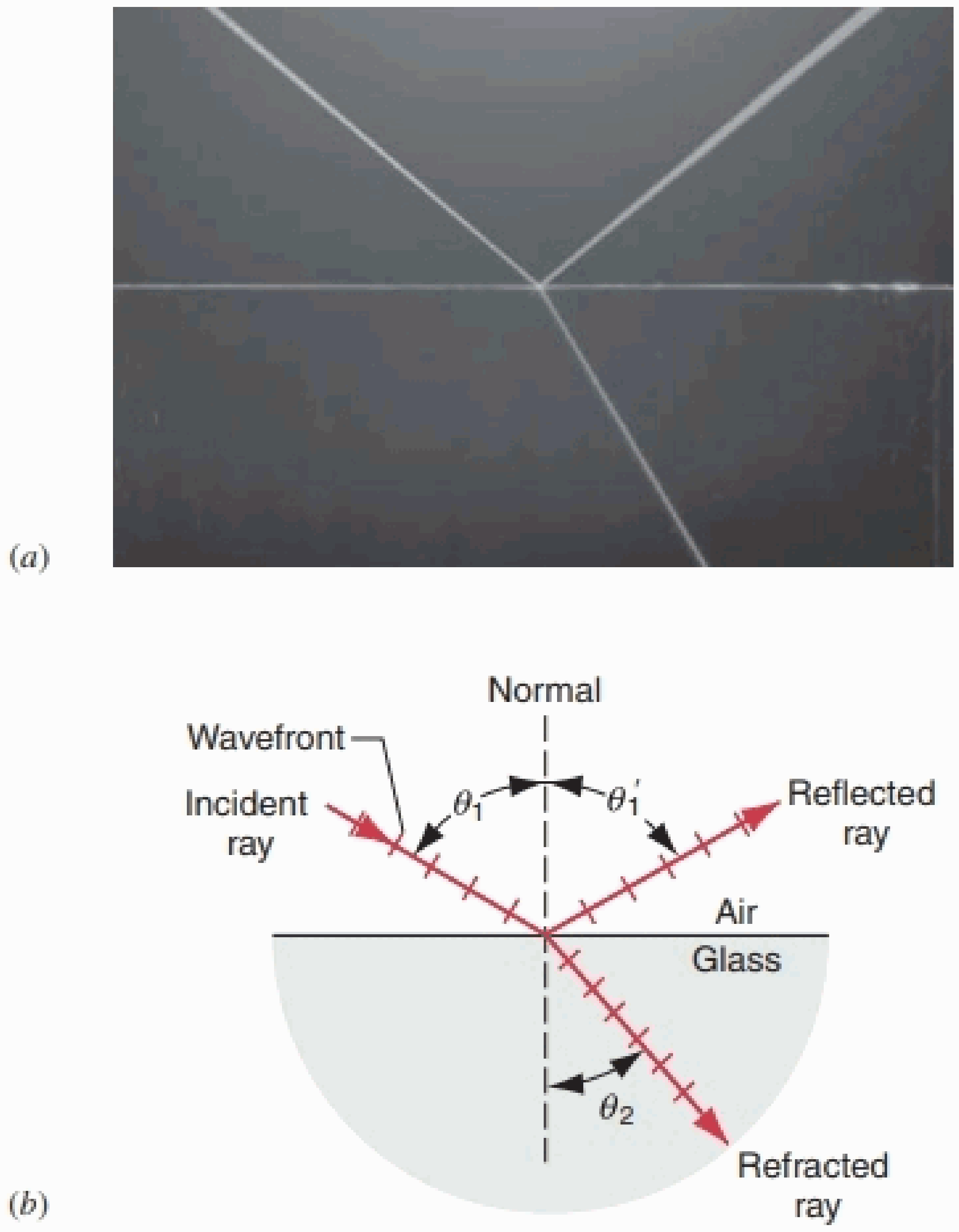


FIGURE 39-10. (*a*) A photograph showing the reflection and refraction of a light beam incident on a plane glass surface. (*b*) A representation using rays. The angles of incidence θ_1, reflection θ_1', and refraction θ_2 are marked. Note that the angles are measured between the normal to the surface and the appropriate ray.

Equation 39-4 is called *Snell's law.* Here n_1 and n_2 are dimensionless constants called the *index of refraction* of medium 1 and medium 2. The index of refraction n of a medium is the ratio between the speed of light c in vacuum and the speed of light v in that medium:

$$n = \frac{c}{v}. \tag{39-5}$$

We discussed the speed of light in various materials in the previous section. It is fair to say that refraction occurs because the speed of light changes from one medium to another. We develop this idea later in this section.

Table 39-3 shows some examples of the index of refraction of various materials. Note that, for most purposes, air

TABLE 39-3 Some Indices of Refraction[a]

Medium	*Index*	*Medium*	*Index*
Vacuum (exactly)	1.00000	Typical crown glass	1.52
Air (STP)	1.00029	Sodium chloride	1.54
Water (20°C)	1.33	Polystyrene	1.55
Acetone	1.36	Carbon disulfide	1.63
Ethyl alcohol	1.36	Heavy flint glass	1.65
Sugar solution (30%)	1.38	Sapphire	1.77
Fused quartz	1.46	Heaviest flint glass	1.89
Sugar solution (80%)	1.49	Diamond	2.42

[a] For a wavelength of 589 nm (yellow sodium light).

**Refracted* comes from the Latin for "broken"; the same root occurs in the word "fracture." If you dip a slanted pencil part way into a bowl of water, the pencil appears to be "broken."

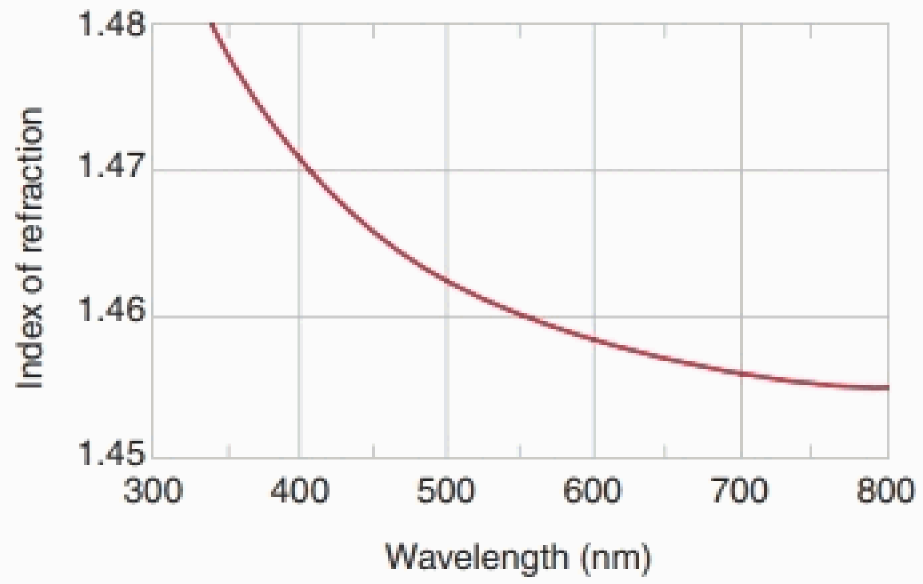


FIGURE 39-11. The index of refraction of fused quartz as a function of wavelength.

can be regarded as equivalent to a vacuum in its refraction of light. The index of refraction of a material generally varies with the wavelength of the light (see Fig. 39-11). Refraction can thus be used to analyze a beam of light into its constituent wavelengths, such as occurs in a rainbow or a glass prism; this effect is called *dispersion.*

Reflection and Refraction of Electromagnetic Waves (Optional)

The laws of reflection and refraction hold for all regions of the electromagnetic spectrum, not only for light. In fact, Eqs. 39-3 and 39-4 can be derived from Maxwell's equations, which makes them generally applicable to electromagnetic waves. Experimental evidence for this general applicability includes the reflection of microwaves or radio waves from the ionosphere and the refraction of x rays by crystals.

We normally think of highly polished or smooth surfaces as "good" reflectors, but other surfaces may reflect as well—for example, a sheet of paper. The reflection by the paper (which is called a *diffuse reflection*) scatters the light more or less in all directions. It is largely by diffuse reflections that we see nonluminous objects around us. The difference between diffuse and specular (mirrorlike) reflection depends on the roughness of the surface: a reflected beam is formed only if the typical dimensions of the surface irregularities of the reflector are substantially less than the wavelength of the incident light. Thus the classification of the reflective properties of a surface depends on the wavelength of the radiation that strikes the surface. The bottom of a cast-iron skillet, for example, may be a good reflector for microwaves of wavelength 0.5 cm but is not a good reflector for visible light.

Maxwell's equations permit us to calculate how the incident energy is divided between the reflected and refracted beams. Figure 39-12 shows the theoretical prediction for (*a*) a light beam in air falling on a glass–air interface, and (*b*) a light beam in glass falling on a glass–air interface. Figure 39-12*a* shows that for angles of incidence up to about 60°, less than 10% of the light energy is reflected. At grazing incidence (that is, at angles of incidence near 90°), the surface becomes an excellent reflector. Another example of this effect is the high reflecting power of a wet road when light from automobile headlights strikes the road near grazing incidence.

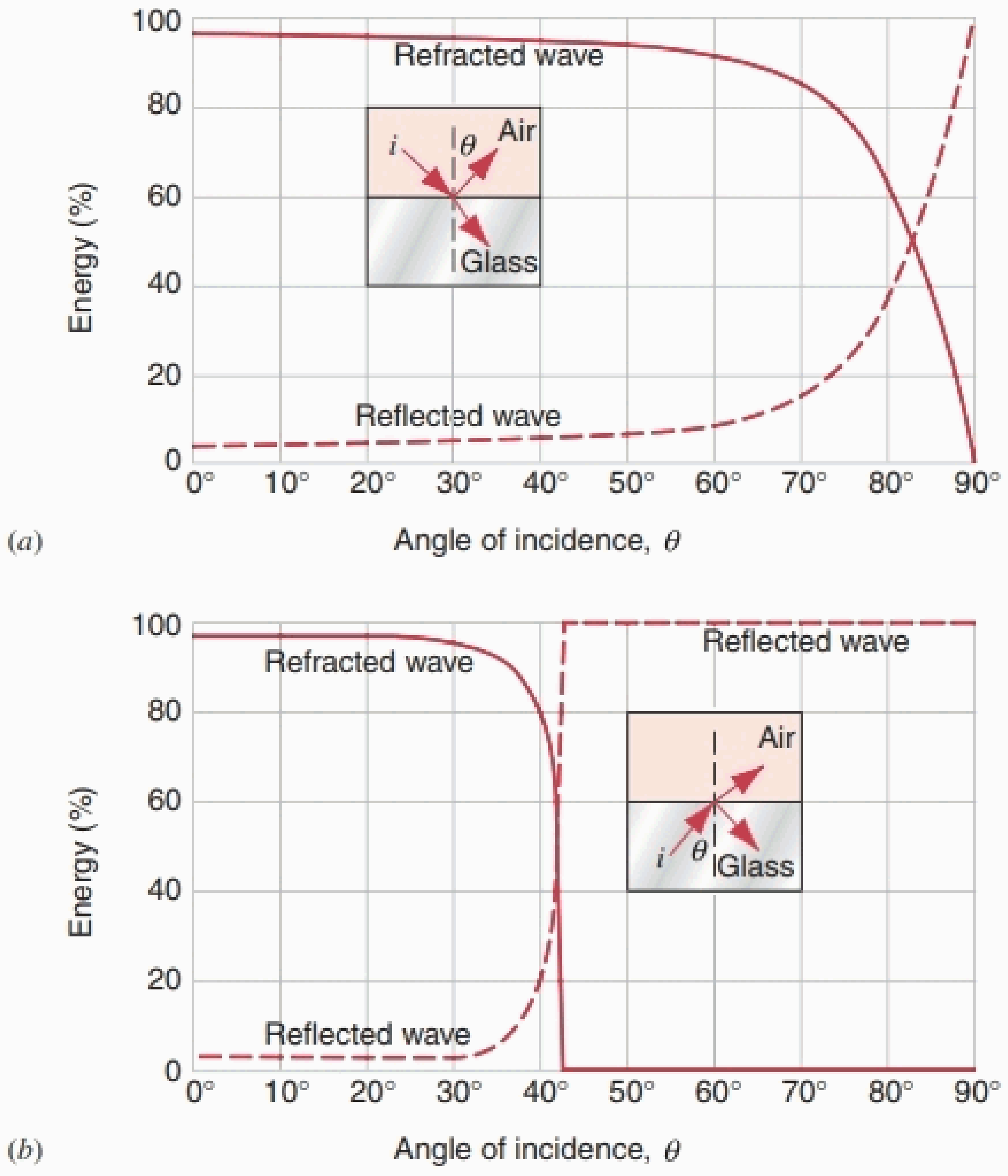


FIGURE 39-12. (*a*) The percentage of energy reflected and refracted when a wave in air is incident on glass ($n = 1.50$). (*b*) The same for a wave in glass incident on air, showing total internal reflection.

Figure 39-12*b* shows clearly that at a certain critical angle (41.8° in this case), *all* the light is reflected. We consider this phenomenon, called *total internal reflection,* in Section 39-5. ■

SAMPLE PROBLEM 39-2. Figure 39-13 shows an incident ray i striking a plane mirror MM' at angle of incidence θ. Mirror $M'M''$ is perpendicular to MM'. Trace this ray through its subsequent reflections.

Solution The reflected ray r makes an angle θ with the normal at b and falls as an incident ray on mirror $M'M''$. Its angle of incidence θ' on this mirror is $\pi/2 - \theta$. A second reflected ray r' makes an angle θ' with the normal erected at b'. Rays i and r' are antiparallel for any value of θ. To see this, note that

$$\phi = \pi - 2\theta' = \pi - 2\left(\frac{\pi}{2} - \theta\right) = 2\theta.$$

Two lines are parallel if their opposite interior angles for an intersecting line (ϕ and 2θ) are equal.

Repeat the problem if the angle between the mirrors is 120° rather than 90°.

The three-dimensional analogue of Fig. 39-13 is the *corner reflector,* which consists of three perpendicular plane mirrors joined like the positive sections of the coordinate planes of an xyz system. A corner reflector has the property that, for *any* direction of incidence, an incident ray is reflected back in the opposite direction. Highway reflectors use this principle, so that light from the headlights of an oncoming car is reflected back toward the car, regardless of the direction of approach of the car or the angle of the headlights above the road. Corner reflectors were placed on the Moon by the Apollo astronauts; timing a reflected laser beam from Earth permits precise determination of the Earth–Moon separation. Corner reflectors can also be found on the masts of small boats, placed there so that search radars can more easily locate them.

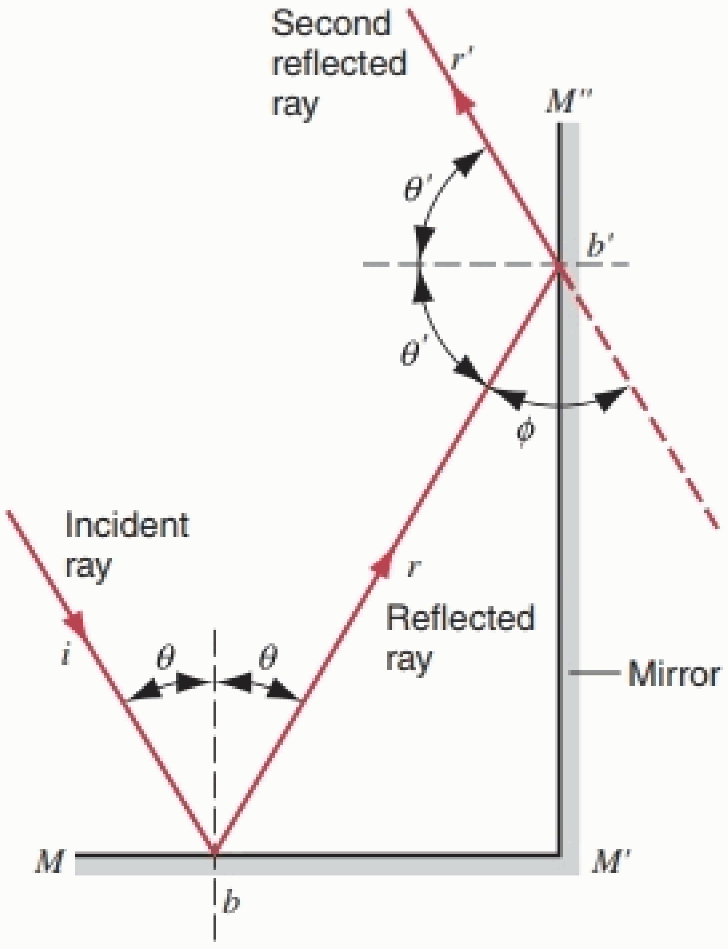


FIGURE 39-13. Sample Problem 39-2. A two-dimensional corner reflector.

SAMPLE PROBLEM 39-3. A light beam in air is incident on the plane surface of a block of quartz and makes an angle of 30.00° with the normal. The beam contains two wavelengths, 400 and 500 nm. The indices of refraction for quartz at these wavelengths are 1.4702 and 1.4624, respectively. What is the angle between the two refracted beams in the quartz?

Solution From Eq. 39-4 ($n_1 \sin\theta_1 = n_2 \sin\theta_2$) we have, for the 400-nm beam (taking $n_1 = 1.0003$ for air)

$$(1.0003)\ \sin 30.00° = (1.4702) \sin\theta_2,$$

which leads to

$$\theta_2 = 19.89°.$$

For the 500-nm beam we have

$$(1.0003)\ \sin 30.00° = (1.4624) \sin\theta_2$$

or

$$\theta_2 = 20.00°.$$

The angle $\Delta\theta_2$ between the beams is 0.11°, with the shorter wavelength component bent through the larger angle—that is, having the smaller angle of refraction. The difference in angle decreases as the angle of incidence decreases, becoming 0.018° for $\theta_1 = 5°$. In optical instruments using lenses, the variation in the refraction angle with wavelength leads to a distortion called *chromatic aberration.* Using small angles of incidence reduces the distortion due to chromatic aberration.

SAMPLE PROBLEM 39-4. A light beam in air is incident on one face of a glass prism as in Fig. 39-14. The angle θ is chosen so that the emerging ray also makes an angle θ with the normal to the other face. Derive an expression for the index of refraction of the prism material, taking $n = 1$ for air.

Solution Note that $\angle\, bad + \alpha = \pi/2$ and that $\angle\, bad + \phi/2 = \pi/2$, where ϕ is the prism angle. Therefore,

$$\alpha = \tfrac{1}{2}\phi. \tag{39-6}$$

The deviation angle ψ is the sum of the two opposite interior angles in triangle *aed,* or

$$\psi = 2(\theta - \alpha).$$

Substituting $\frac{1}{2}\phi$ for α and solving for θ yield

$$\theta = \tfrac{1}{2}(\psi + \phi). \tag{39-7}$$

At point a, θ is the angle of incidence and α is the angle of refraction. The law of refraction (see Eq. 39-4) is

$$\sin\theta = n \sin\alpha,$$

in which n is the index of refraction of the glass.

From Eqs. 39-6 and 39-7 this yields

$$\sin\frac{\psi + \phi}{2} = n \sin\frac{\phi}{2}$$

or

$$n = \frac{\sin(\psi + \phi)/2}{\sin(\phi/2)}, \tag{39-8}$$

which is the desired relation. This equation holds only for θ chosen so that the light ray passes symmetrically through the prism.

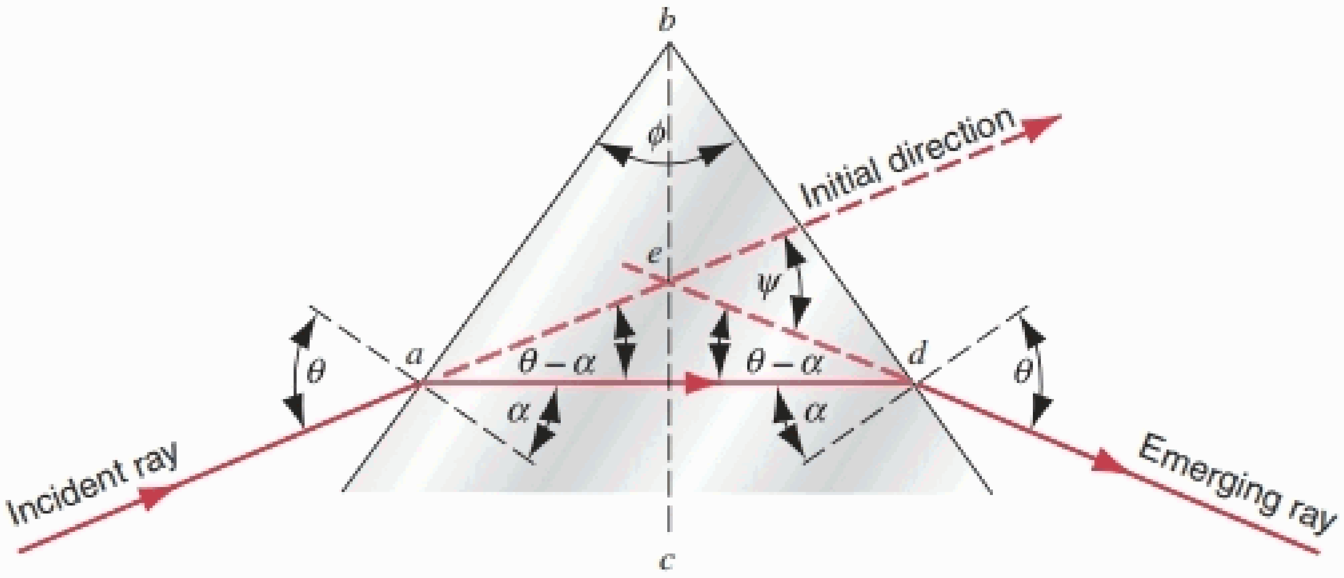


FIGURE 39-14. Sample Problem 39-4.

In this case ψ is called the *angle of minimum deviation*; if θ is either increased or decreased, a larger deviation angle results.

Deriving the Law of Reflection

The law of reflection can be derived in several different ways. We discuss two of these derivations here.

Huygens' Principle. The Dutch physicist Christiaan Huygens* put forth a simple theory of light in 1678. This theory assumes that light is a wave, but it says nothing about the nature of the wave. (In particular, since Maxwell's theory of electromagnetism would not appear for nearly two centuries, Huygens' theory gives no hint of the electromagnetic character of light.) Huygens did not know whether light was a transverse wave or a longitudinal one; he did not know the wavelengths of visible light; he had little knowledge of the speed of light. Nevertheless, his theory was a useful guide to experiment for many years and remains useful today for pedagogic and certain other practical purposes. We must not expect it to yield the same wealth of detailed information that Maxwell's more complete electromagnetic theory does.

Huygens' theory is based on a geometrical construction that allows us to tell where a given wavefront will be at any time in the future if we know its present position. Huygens' principle can be stated as follows:

> *All points on a wavefront can be considered as point sources for the production of spherical secondary wavelets. After a time t the new position of a wavefront is the surface tangent to these secondary wavelets.*

Consider a trivial example. Given a wavefront (*ab* in Fig. 39-15) in a plane wave in free space, where will the wavefront be a time t later? Following Huygens' principle, we let several points on this plane (the dots in Fig. 39-15) serve as centers for secondary spherical wavelets. In a time t the radius of these spherical waves is ct, where c is the speed of light in free space. We represent the plane tangent to these spheres at time t by *de*. As we expect, it is parallel to plane *ab* and a perpendicular distance ct from it. Thus plane wavefronts are propagated as planes and with speed c. Note that the Huygens method involves a three-dimensional construction and that Fig. 39-15 is the intersection of this construction with the plane of the page.

We might expect that, contrary to observation, a wave should be radiated backward as well as forward from the dots in Fig. 39-15. This result is avoided by assuming that the intensity of the spherical wavelets is not uniform in all directions but varies continuously from a maximum in the forward direction to a minimum of zero in the back direction. This is suggested by the shading of the circular arcs in Fig. 39-15. Huygens' method can be applied quantitatively

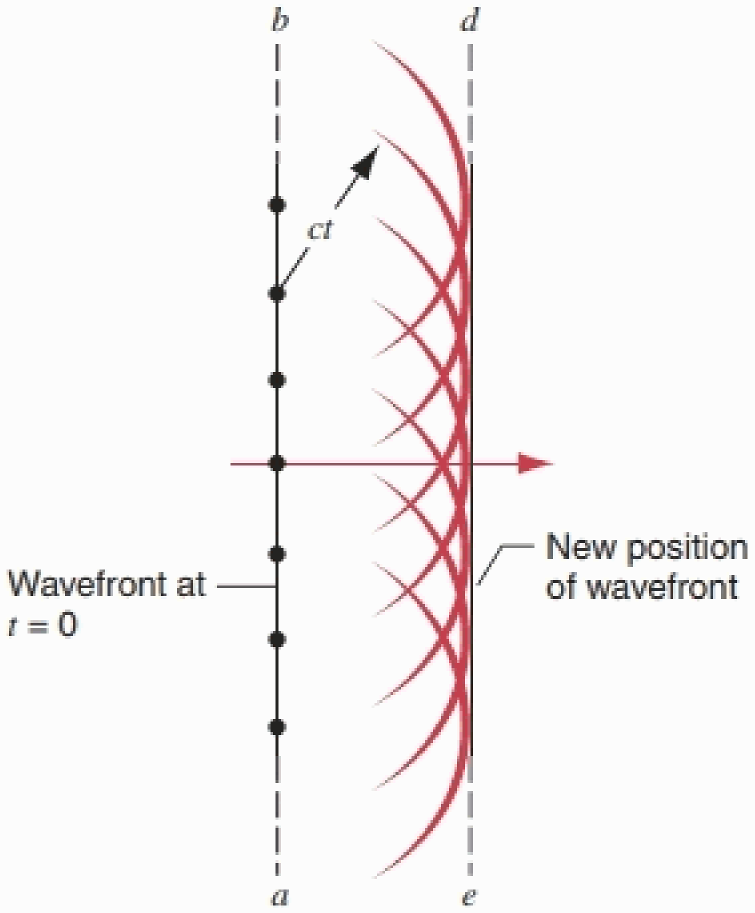


FIGURE 39-15. The propagation of a plane wave in free space is described by the Huygens construction. Note that the ray (horizontal arrow) representing the wave is perpendicular to the wavefronts.

---

*Christiaan Huygens (1629–1695) was a scientist of remarkable depth and influence. In addition to the wave theory of light, his accomplishments include improvements in telescope design that permitted him to deduce the shape of the rings of Saturn, the development of the pendulum clock, and contributions to the theory of rotating bodies (including the first recognition of the existence of centripetal acceleration) and colliding objects (including the principle of conservation of momentum).

to *all* wave phenomena; see Exercise 12. The method was put on a firm mathematical footing two centuries after Huygens by Gustav Kirchhoff (1824–1887), who proved that the intensity of the wavelets varies with direction as described above.

Now we show how the law of reflection follows from Huygens' principle. Figure 39-16*a* shows three wavefronts in a plane wave falling on a plane mirror MM'. For convenience the wavefronts are chosen to be one wavelength apart. Note that θ_1, the angle between the wavefronts and the mirror, is the same as the angle between the incident ray and the normal to the mirror. In other words, θ_1 is the *angle of incidence.* The three wavefronts are related to each other by the Huygens construction, as in Fig. 39-15.

Let us regard point a in the wavefront in Fig. 39-16*b* as a source of a Huygens wavelet, which expands after a time λ/c to include point b on the surface of the mirror. Light from point p in this same wavefront cannot move beyond the mirror but must expand upward as a spherical Huygens wavelet. Setting a compass to radius λ and swinging an arc about p provides a semicircle to which the reflected wavefront must be tangent. Since point b must lie on the new wavefront, this tangent must pass through b. Note that the angle θ_1' between the wavefront and the mirror is the same as the angle between the reflected ray and the normal to the mirror. In other words, θ_1' is the *angle of reflection.*

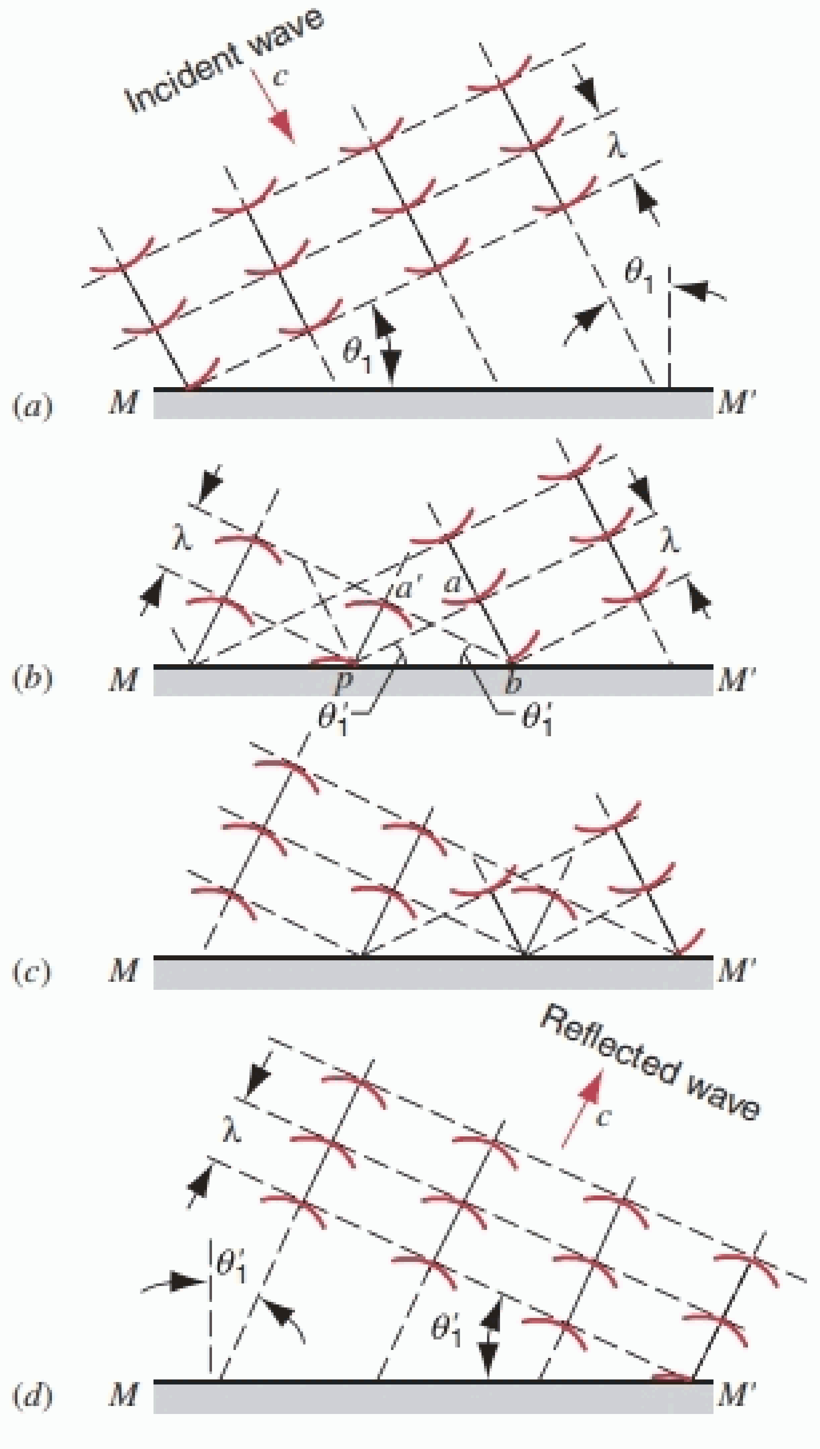


FIGURE 39-16. The reflection of a plane wave from a plane mirror as analyzed by the Huygens construction.

Consider right triangles abp and $a'bp$. They have side bp in common, and side ab $(=\lambda)$ is equal to side $a'p$. The two right triangles are thus congruent and we may conclude that

$$\theta_1 = \theta_1',$$

verifying the law of reflection. If you recall that the Huygens construction is three dimensional and that the arcs shown represent segments of spherical surfaces, you will be able to convince yourself that the reflected ray lies in the plane formed by the incident ray and the normal to the mirror—that is, the plane of Fig. 39-16. This is also a requirement of the law of reflection.

Figures 39-16*c* and 39-16*d* show how the process continues until all three incident wavefronts have been reflected.

Fermat's Principle. In 1650, Pierre Fermat* discovered a remarkable principle, which we can express in these terms:

> *A light ray traveling from one fixed point to another fixed point follows a path such that, compared with nearby paths, the time required is either a minimum or a maximum or remains unchanged (that is, stationary).*

We can readily derive the law of reflection from this principle. Figure 39-17 shows two fixed points A and B and a reflecting ray APB connecting them. (We assume that ray APB lies in the plane of the figure; see Exercise 11.) The total length L of this ray is

$$L = \sqrt{a^2 + x^2} + \sqrt{b^2 + (d - x)^2},$$

where x locates the point P at which the ray touches the mirror.

According to Fermat's principle, P will have a position such that the time of travel $t = L/c$ of the light must be a minimum (or a maximum or must remain unchanged), which occurs when $dt/dx = 0$. Taking this derivative yields

$$\begin{aligned}\frac{dt}{dx} &= \frac{1}{c}\frac{dL}{dx}\\ &= \frac{1}{2c}(a^2 + x^2)^{-1/2}(2x)\\ &\quad + \frac{1}{2c}[b^2 + (d - x)^2]^{-1/2}(2)(d - x)(-1) = 0,\end{aligned}$$

---

*Pierre Fermat (1601–1665) was a French mathematician who is remembered for his development of analytic geometry and his many contributions to number theory. Perhaps his most challenging result is known as *Fermat's last theorem:* the equation $x^n + y^n = z^n$, in which x, y, z, and n are positive integers, has no solution for $n > 2$. Despite Fermat's claim of a proof of this theorem (a proof that he failed to publish), it eluded mathematicians for more than 300 years, until it was solved in 1994 by Princeton mathematician Andrew Wiles.

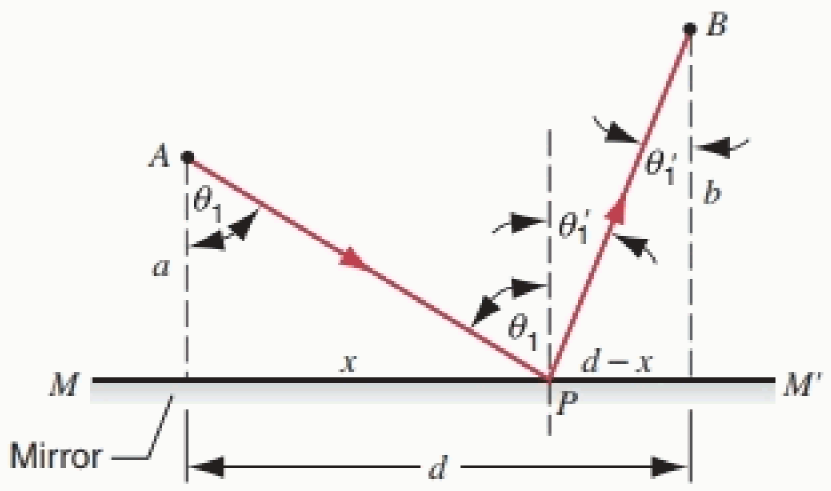


FIGURE 39-17. The reflection of a plane wave from a plane mirror as analyzed using Fermat's principle. A ray from A passes through B after reflection at P.

which we can rewrite as

$$\frac{x}{\sqrt{a^2+x^2}}=\frac{d-x}{\sqrt{b^2+(d-x)^2}}.$$

(In evaluating the derivative, note that we hold the endpoints fixed and vary the path by allowing x to vary.) Comparison with Fig. 39-17 shows that we can rewrite this as

$$\sin\theta_1=\sin\theta_1'$$

or

$$\theta_1=\theta_1',$$

which is the law of reflection.

Deriving the Law of Refraction

Now we use Huygens' principle and Fermat's principle to derive the law of refraction (Eq. 39-4).

Huygens' Principle. Figure 39-18 shows four stages in the refraction of three successive wavefronts in a plane wave falling on an interface between air (medium 1) and glass (medium 2). For convenience, we assume that the incident wavefronts are separated by λ_1, the wavelength as measured in medium 1. Let the speed of light in air be v_1 and that in glass be v_2. We assume that

$$v_2 < v_1. \tag{39-9}$$

This assumption about the speeds is vital to the derivation that follows.

The wavefronts in Fig. 39-18a are related to each other by the Huygens construction of Fig. 39-15. As in Fig. 39-16, θ_1 is the angle of incidence. In Fig. 39-18b consider the time $(=\lambda_1/v_1)$ during which a Huygens wavelet from point e moves to include point d. Light from point h, traveling through glass at a reduced speed (recall the assumption of Eq. 39-9) moves a shorter distance,

$$\lambda_2=\lambda_1\frac{v_2}{v_1}, \tag{39-10}$$

during this time. This follows from $v=\lambda f$ and $f_1=f_2$. The refracted wavefront must be tangent to an arc of this radius centered on h. Since d lies on the new wavefront, the tangent must pass through this point, as shown. Note that θ_2, the angle between the refracted wavefront and the air–glass interface, is the same as the angle between the refracted ray and the normal to this interface. In other words, θ_2 is the *angle of refraction*. Note too that the wavelength in glass (λ_2) is less than the wavelength in air (λ_1).

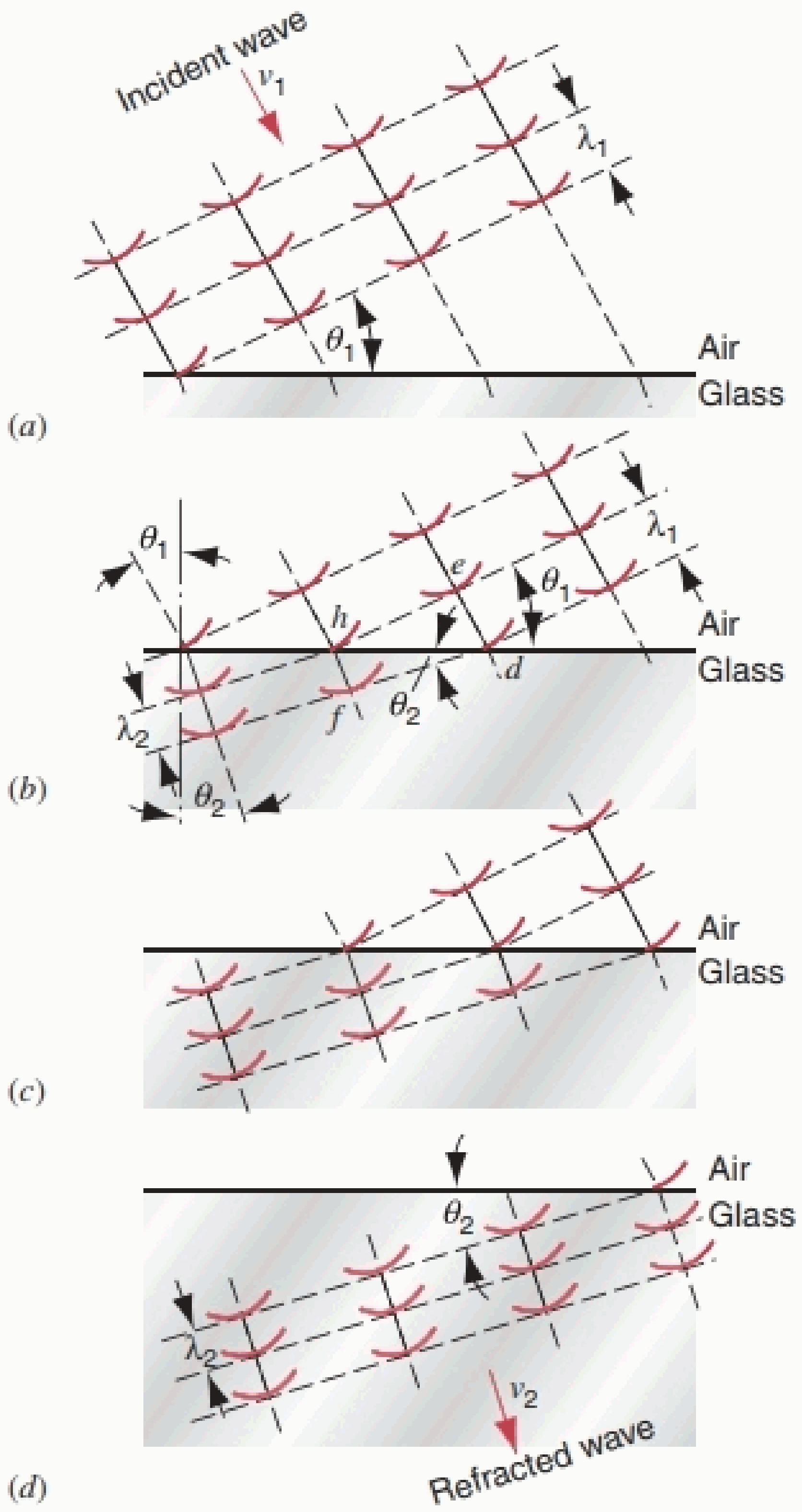


FIGURE 39-18. The refraction of a plane wave at a plane interface as described by the Huygens construction. For simplicity the reflected wave is not shown. Note the change in wavelength of the refracted wave.

For the right triangles hde and hdf we may write

$$\sin\theta_1=\frac{\lambda_1}{hd}\qquad(\text{for } hde)$$

and

$$\sin\theta_2=\frac{\lambda_2}{hd}\qquad(\text{for } hdf).$$

Dividing and using Eq. 39-10 yields

$$\frac{\sin\theta_1}{\sin\theta_2}=\frac{\lambda_1}{\lambda_2}=\frac{v_1}{v_2}. \tag{39-11}$$

Introducing a common factor of c allows us to rewrite Eq. 39-11 as

$$\frac{c}{v_1}\sin\theta_1 = \frac{c}{v_2}\sin\theta_2. \qquad (39\text{-}12)$$

According to Eq. 39-5, $c/v_1 = n_1$ and $c/v_2 = n_2$, so that Eq. 39-12 becomes

$$n_1 \sin\theta_1 = n_2 \sin\theta_2, \qquad (39\text{-}13)$$

which is the law of refraction.

If one medium is a vacuum, Eq. 39-10 becomes

$$\lambda_n = \lambda\frac{v}{c} = \frac{\lambda}{n}, \qquad (39\text{-}14)$$

where λ_n is the wavelength of light in a medium of index n and λ is the wavelength in vacuum. In passing from one medium to another, both the speed of light and its wavelength are reduced by the same factor, but the frequency of the light is unchanged.

The application of Huygens' principle to refraction requires that if a light ray is bent toward the normal in passing from air to an optically dense medium, then the speed of light in that optically dense medium (glass, say) must be *less* than that in air; see Eq. 39-9. This requirement holds for all wave theories of light. In the early particle theory of light put forward by Newton, the explanation of refraction required that the speed of light in the medium in which light is bent toward the normal (the optically denser medium) be *greater* than that in air. The denser medium was thought to exert attractive forces on the light "corpuscles" as they neared the surface, speeding them up and changing their direction to cause them to make a smaller angle with the normal.

An experimental comparison of the speed of light in air and in an optically denser medium is therefore critical in deciding between the wave and corpuscular theories of light. Such a measurement was first carried out by Foucault in 1850; he showed conclusively that light travels more slowly in water than in air, thus ruling out the corpuscular theory of Newton.

Fermat's Principle. To prove the law of refraction from Fermat's principle, consider Fig. 39-19, which shows two fixed points A and B in two different media and a refracting ray APB connecting them. The time t for the ray to travel from A to B is given by

$$t = \frac{L_1}{v_1} + \frac{L_2}{v_2}.$$

Using the relation $n = c/v$, we can write this as

$$t = \frac{n_1L_1 + n_2L_2}{c} = \frac{L}{c}, \qquad (39\text{-}15)$$

where L is the *optical path length,* defined as

$$L = n_1L_1 + n_2L_2. \qquad (39\text{-}16)$$

For any light ray traveling through successive media, the optical path length is the sum of the products of the geometrical path length of each segment and the index of refraction of that medium. Equation 39-14 ($\lambda_n = \lambda/n$) shows that the optical path length is equal to the length that this same number of waves would have if the medium were a vacuum. Do not confuse the optical path length with the geometrical path length, which is $L_1 + L_2$ for the ray of Fig. 39-19.

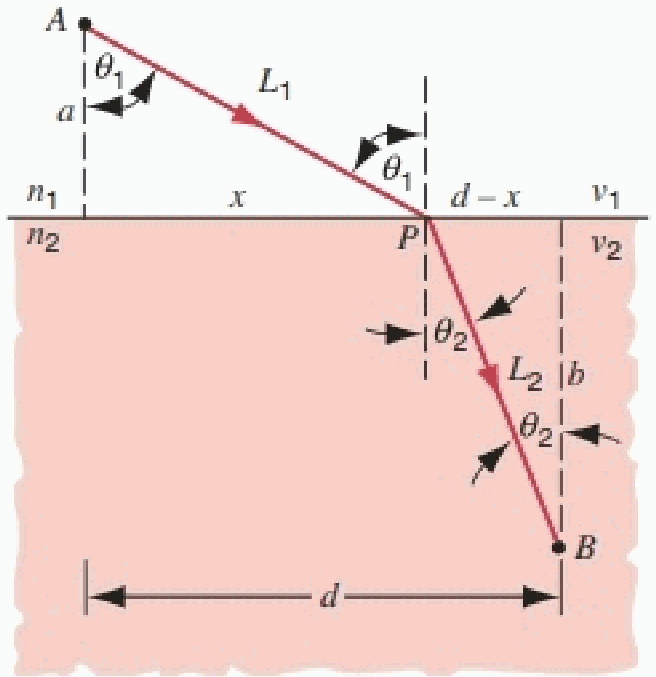


FIGURE 39-19. The refraction of a plane wave at a plane interface as analyzed using Fermat's principle. A ray from A passes through B after refraction at P.

Fermat's principle requires that the time t for the light to travel the path APB must be a minimum (or a maximum or must remain unchanged), which in turn requires that x be chosen so that $dt/dx = 0$. The optical path length in Fig. 39-19 is

$$L = n_1L_1 + n_2L_2 = n_1\sqrt{a^2 + x^2} + n_2\sqrt{b^2 + (d - x)^2}.$$

Substituting this result into Eq. 39-15 and differentiating, we obtain

$$\begin{aligned}\frac{dt}{dx} &= \frac{1}{c}\frac{dL}{dx}\\ &= \frac{n_1}{2c}(a^2 + x^2)^{-1/2}(2x)\\ &\quad + \frac{n_2}{2c}[b^2 + (d - x)^2]^{-1/2}(2)(d - x)(-1) = 0,\end{aligned}$$

which we can write as

$$n_1\frac{x}{\sqrt{a^2 + x^2}} = n_2\frac{d - x}{\sqrt{b^2 + (d - x)^2}}.$$

Comparison with Fig. 39-19 shows that we can write this as

$$n_1 \sin\theta_1 = n_2 \sin\theta_2,$$

which is the law of refraction.

SAMPLE PROBLEM 39-5. Red light of wavelength 632 nm in free space is incident, at an angle of $\theta_1 = 39°$ with respect to the normal, on a glass microscope slide of thickness $d = 0.78$ mm and index of refraction $n = 1.52$ (Fig. 39-20). Find (*a*) the wavelength in the glass and (*b*) the optical path length of the light in traveling through the glass.

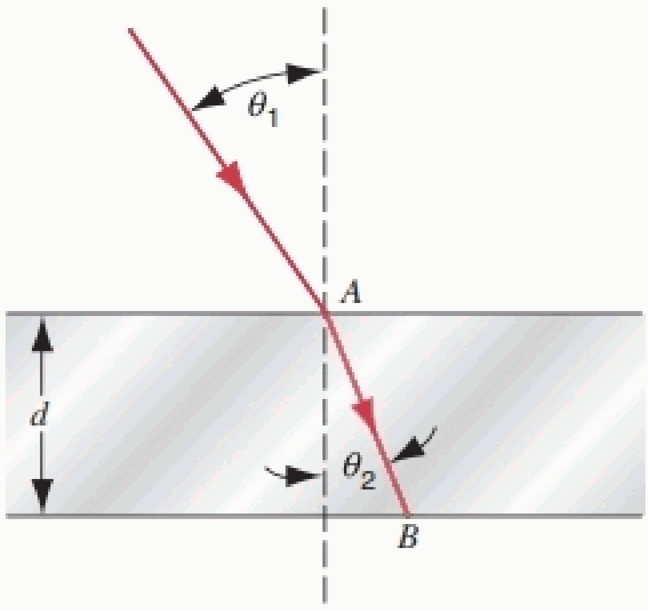


FIGURE 39-20. Sample Problem 39-5.

Solution (*a*) We can find the wavelength in the glass using Eq. 39-14, which gives

$$\lambda_n = \frac{\lambda}{n} = \frac{632 \text{ nm}}{1.52} = 416 \text{ nm}.$$

(*b*) The angle of refraction is found from Eq. 39-13,

$$\sin\theta_2 = \frac{\sin\theta_1}{n} = \frac{\sin 39^\circ}{1.52} = 0.414$$

or

$$\theta_2 = 24.5^\circ,$$

and the actual length of the path through the glass is

$$AB = \frac{d}{\cos\theta_2} = \frac{0.78 \text{ mm}}{\cos 24.5^\circ} = 0.857 \text{ mm}.$$

The optical path is

$$L = n(AB) = 1.52(0.857 \text{ mm}) = 1.30 \text{ mm}.$$

39-5 TOTAL INTERNAL REFLECTION

Figure 39-21 shows rays from a point source in glass falling on a glass–air interface. As the angle of incidence θ is increased, we reach a situation (see ray *e*) at which the refracted ray points along the surface, the angle of refraction being 90°. For angles of incidence larger than this *critical angle* θ_c, there is no refracted ray, and we speak of *total internal reflection.*

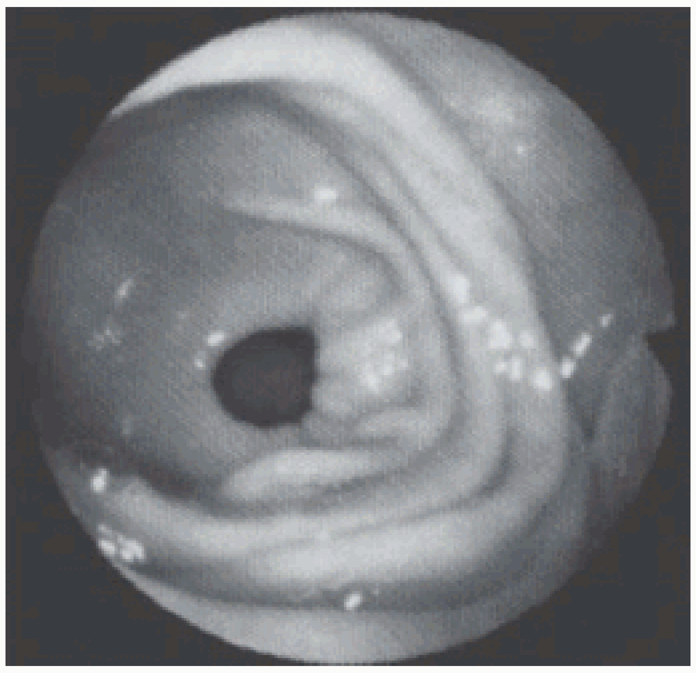


FIGURE 39-22. Fiber optic image of the passage from the stomach into the small intestine.

We find the critical angle by putting $\theta_2 = 90^\circ$ in the law of refraction (see Eq. 39-4):

$$n_1 \sin\theta_c = n_2 \sin 90^\circ$$

or

$$\theta_c = \sin^{-1}\frac{n_2}{n_1}. \qquad (39\text{-}17)$$

For glass in air, $\theta_c = \sin^{-1}(1.00/1.50) = 41.8^\circ$. Figure 39-12b indicates that 100% of the energy of the wave is reflected when the angle of incidence exceeds 41.8°.

The sine of an angle cannot exceed unity, so we must have $n_2 < n_1$. This tells us that total internal reflection cannot occur when the incident light is in the medium of lower index of refraction. The word *total* means just that; the reflection occurs with no loss of intensity. In ordinary reflection from a mirror, by way of contrast, there is an intensity loss of about 4%.

Total internal reflection makes possible fiber optical devices that physicians can use to visually inspect many internal body sites; see Fig. 39-22. In these devices, a bundle of fibers transmits an image that can be inspected visually outside the body.* Optical fibers are also used for telephone communications, for electronic data transmission,

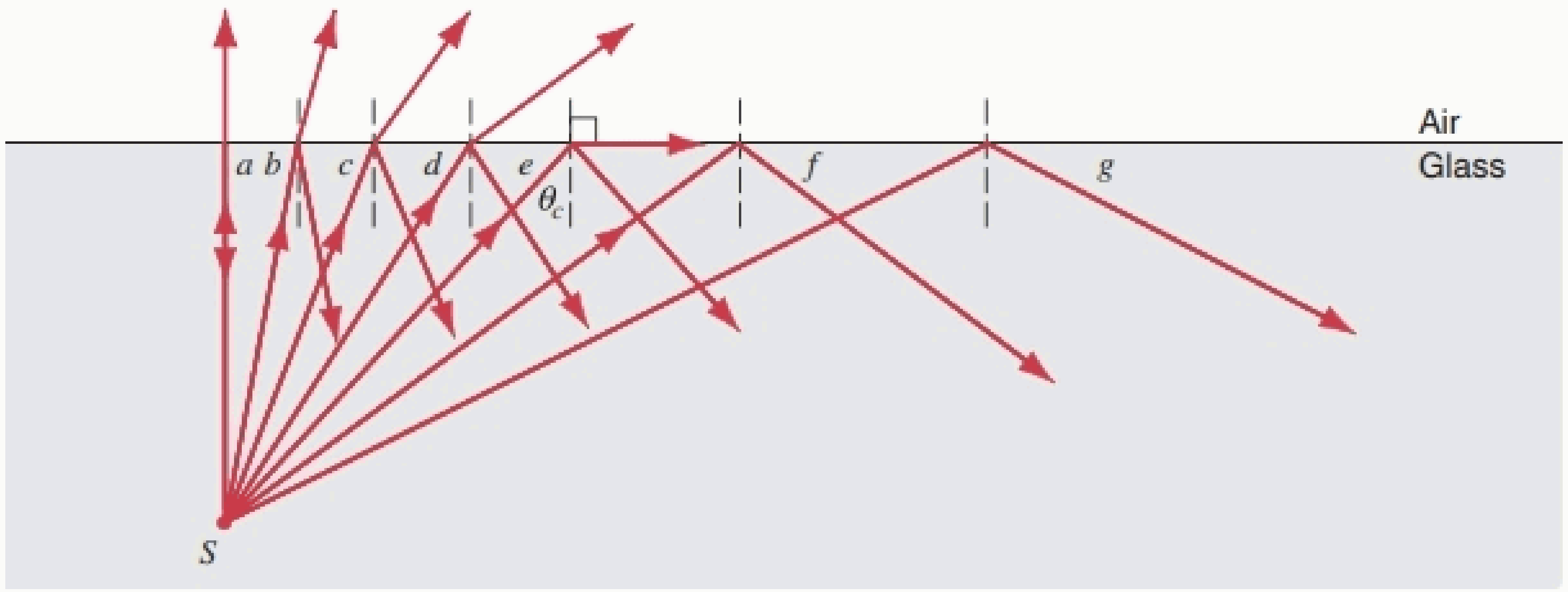


FIGURE 39-21. Total internal reflection of light from a point source S occurs for all angles of incidence greater than the critical angle θ_c. At the critical angle, the refracted ray points along the air–glass interface.

---

*See "Optical Fibers in Medicine," by Abraham Katzir, *Scientific American,* May 1989, p. 120.

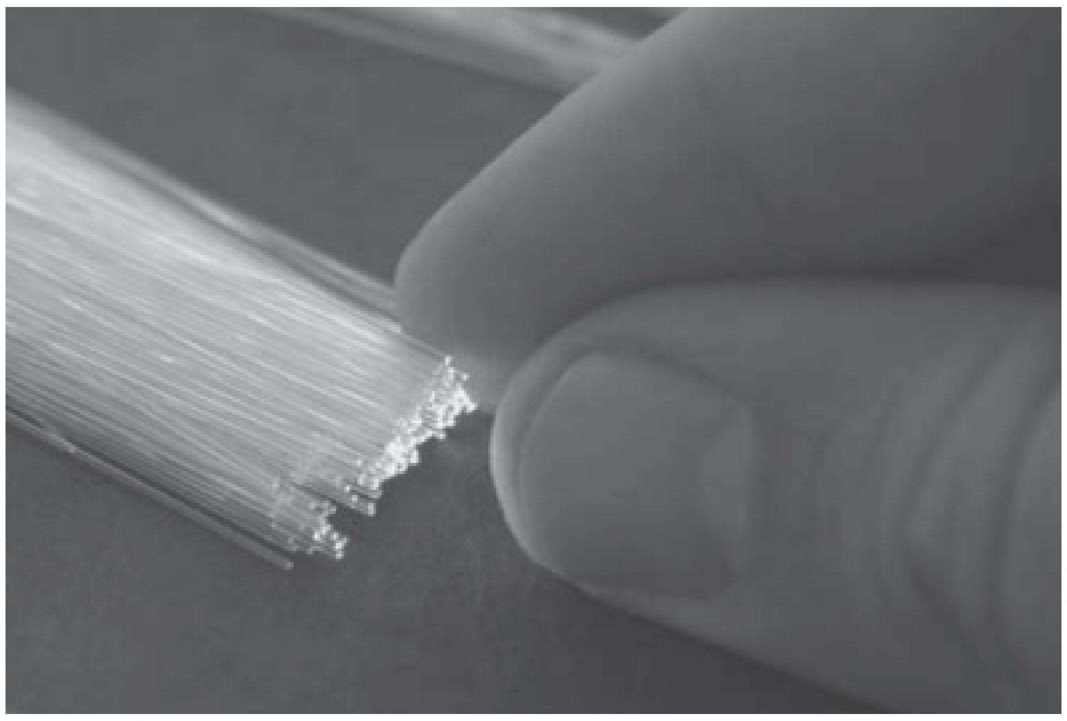


FIGURE 39-23. Light is transmitted through an optical fiber.

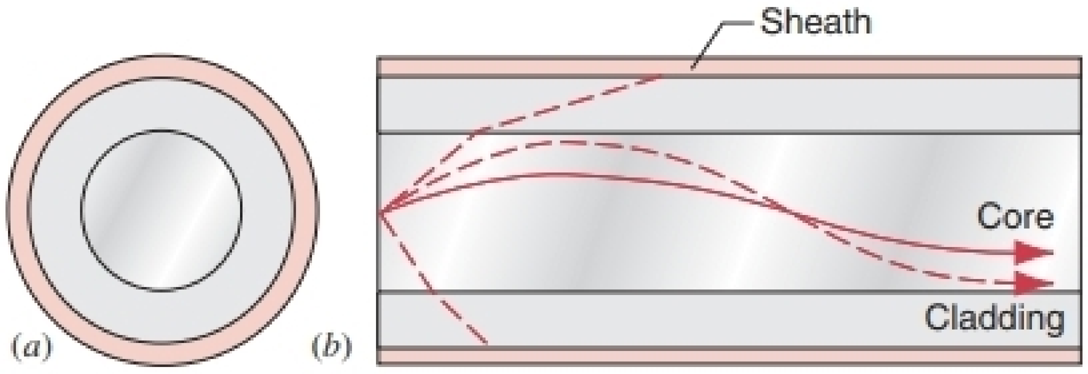


FIGURE 39-24. (*a*) An optical fiber shown in cross section. The diameter of the fiber is about the same as that of a human hair. (*b*) A transverse view, showing propagation by total internal reflection. The core, the cladding (of lower index than the core), and the protective sheath are shown.

and, because of their light weight and freedom from electromagnetic interference, for carrying signals on aircraft. Figure 39-23 shows light emerging from an optical fiber.*

As Fig. 39-24 shows, the fiber consists of a central core that is graded smoothly into an outer cladding layer of a material of lower index of refraction. Only those rays that are internally reflected can be propagated along the fiber. To reduce attenuation of the signal as it passes along the fiber, materials of extreme purity have been developed. If seawater were as transparent as the glass from which optical fibers are made, it would be possible to see the sea bottom by reflected sunlight at a depth of several miles.

SAMPLE PROBLEM 39-6. Figure 39-25*a* shows a triangular prism of glass, a ray incident normal to one face being totally reflected. If θ_1 is 45°, what can you conclude about the index of refraction n of the glass?

Solution The angle θ_1 must be equal to or greater than the critical angle θ_c, where θ_c is given by Eq. 39-17:

$$\theta_c = \sin^{-1}\frac{n_2}{n_1} = \sin^{-1}\frac{1}{n}$$

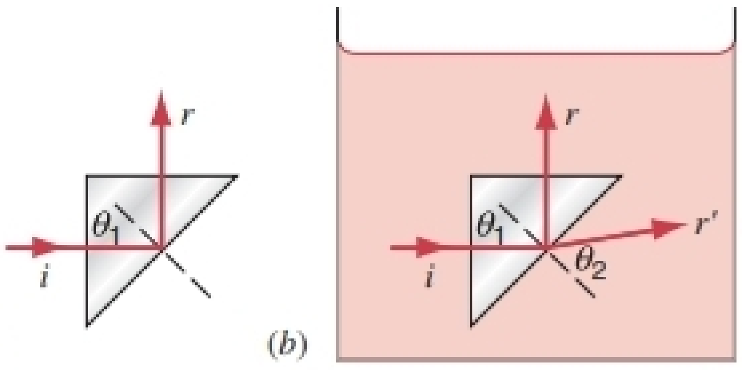


FIGURE 39-25. (*a*) Sample Problem 39-6. (*b*) Sample Problem 39-7.

or

$$n = \frac{1}{\sin\theta_c},$$

in which the index of refraction of air $(= n_2)$ is set equal to unity. Since total internal reflection occurs, θ_c must be less than 45°, and so

$$n > \frac{1}{\sin 45^\circ} = 1.41.$$

Thus the index of refraction of the glass must exceed 1.41. If n were less than 1.41, the ray shown in Fig. 39-25*a* would be partially refracted into the air, instead of totally reflected back into the glass.

SAMPLE PROBLEM 39-7. What happens if the prism in Sample Problem 39-6 (assume that $n_1 = 1.50$) is immersed in water ($n_2 = 1.33$)? See Fig. 39-25*b*.

Solution The new critical angle, given by Eq. 39-17, is

$$\theta_c = \sin^{-1}\frac{n_2}{n_1} = \sin^{-1}\frac{1.33}{1.50} = 62.5^\circ.$$

The actual angle of incidence $(= 45^\circ)$ is less than this, so we do *not* have total internal reflection.

There is a reflected ray r, with an angle of reflection of 45°, as Fig. 39-25*b* shows. There is also a refracted ray r', with an angle of refraction given by

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$
$$(1.50)(\sin\ 45^\circ) = (1.33)\sin\theta_2,$$

which yields $\theta_2 = 52.9^\circ$. Show that as $n_2 \rightarrow n_1$, $\theta_c \rightarrow 90^\circ$.

How does the incident ray i in Figs. 39-25*a* and 39-25*b* determine whether there is air or water beyond the glass? That is, how does it "know" whether to be totally reflected or partially refracted? The traveling wave in glass establishes electric and magnetic fields that are strongly decreasing exponential functions of distance that penetrate a few wavelengths into the next medium. These fields are not those associated with a traveling wave, but can be regarded as "sampling" the medium beyond the boundary. We can demonstrate this penetration by placing a second glass prism near the first, as in Fig. 39-26. In sampling medium 2 (the air), the fields also sense the second prism; although waves are forbidden by the law of refraction from appearing in the air gap between the prisms, they are not forbidden from propagating in the second prism. Note from Fig.

---

*See "Lightwaves and Telecommunication," by Stewart E. Miller, *American Scientist,* January–February 1984, p. 66, and "Light-Wave Communications," by W. S. Boyle, *Scientific American,* August 1977, p. 40.

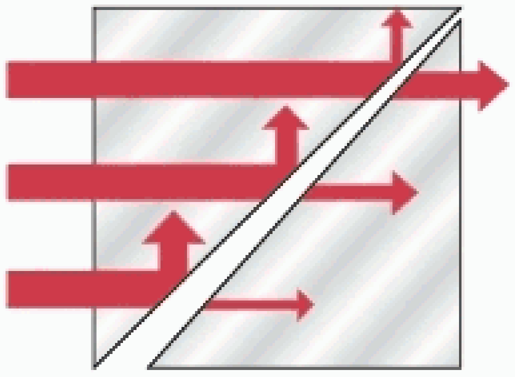


FIGURE 39-26. Frustrated total internal reflection. The thicker the air gap is, the smaller is the intensity of the light in the second prism (indicated by the width of the rays). Note that light waves do not appear *in* the gap.

39-26 that the light ray appears in the second prism but not in the air gap. This situation is called *frustrated total internal reflection* and is a general property of waves. (It can be done, for instance, with microwaves.) In quantum mechanics the wavelike properties of material particles permit a similar effect called *barrier penetration:* a particle can pass from one allowed region to another allowed region by penetrating a region in which it is forbidden. We consider barrier penetration in Chapter 46.

39-6 THE DOPPLER EFFECT FOR LIGHT

In Section 19-9, we showed that if a source of *sound* is moving toward an observer at a speed u, the frequency f heard by the observer is (see Eq. 19-42, which we have rearranged and in which we have substituted u for v_S)

$$f = f_0 \frac{1}{1 + u/v} \qquad \text{(sound wave, observer fixed, source moving away).} \qquad (39\text{-}18)$$

In this equation f_0 is the frequency heard when the source is at rest, and v is the speed of sound. This change in frequency due to relative motion is called the *Doppler effect.*

If the source is at rest in the transmitting medium but the observer is moving away from the source at speed u, the observed frequency (see Eq. 19-39, in which u has been substituted for v_O) is

$$f = f_0(1 - u/v) \qquad \text{(sound wave, source fixed, observer moving away).} \qquad (39\text{-}19)$$

For identical values of the relative separation speed u of the source and the observer, the frequencies predicted by Eqs. 39-18 and 39-19 are different. This is not surprising, because a source of sound moving through a medium in which the observer is at rest is physically different from an observer moving through that medium with the source at rest, as we see by comparing Figs. 19-14 and 19-15 and as we demonstrated in Sample Problem 19-6.

We might be tempted to apply Eqs. 39-18 and 39-19 to light, substituting c, the speed of light, for v, the speed of sound. For light, as contrasted with sound, however, it has proved impossible to identify a medium of transmission relative to which the source and the observer are moving. This means that "source moving away from observer" and "observer moving away from source" are physically identical situations and must exhibit *exactly the same* Doppler-shifted frequency. As applied to light, either Eq. 39-18 or Eq. 39-19 or both must be incorrect. As we show later in this section, the Doppler effect predicted by the theory of relativity is

$$f = f_0 \frac{1 - u/c}{\sqrt{1 - u^2/c^2}} \qquad \text{(light wave, source and observer separating)} \qquad (39\text{-}20)$$

Equation 39-20 applies only to the special case in which the direction of propagation of the light is the same as the direction of the relative motion of S and S'. We can modify Eq. 39-20 if the source and the observer are approaching each other by replacing u with $-u$.

Using the binomial theorem, we can expand the square root in Eq. 39-20 as $(1 - u^2/c^2)^{-1/2} = 1 + \frac{1}{2}u^2/c^2 + \cdots$, which gives

$$f = f_0\left[1 - \frac{u}{c} + \frac{1}{2}\left(\frac{u}{c}\right)^2 + \cdots\right]. \qquad (39\text{-}21)$$

The ratio u/c for most light sources, even those of atomic dimensions, is small. The terms in u^2/c^2 (and higher-order terms) in such cases are negligibly small, and the first-order term u/c gives a reasonable estimate of the Doppler shift.

The classic experimental test of the relativistic Doppler effect was done in 1938 by H. E. Ives and G. R. Stilwell. They sent a beam of hydrogen atoms, generated in a gas discharge, down a tube at speed u, as in Fig. 39-27. They could observe light emitted by these atoms in a direction opposite to $\vec{\mathbf{u}}$ (atom 1, for example) using a mirror, and also in a direction parallel to $\vec{\mathbf{u}}$ (atom 2, for example). With a precise spectrograph, they could photograph a particular characteristic spectral line of this light, obtaining, on a frequency scale, the lines marked f_1' and f_2' in Fig. 39-27*b*. It is also possible to photograph, on the same photographic plate, a line corresponding to light emitted from atoms *at*

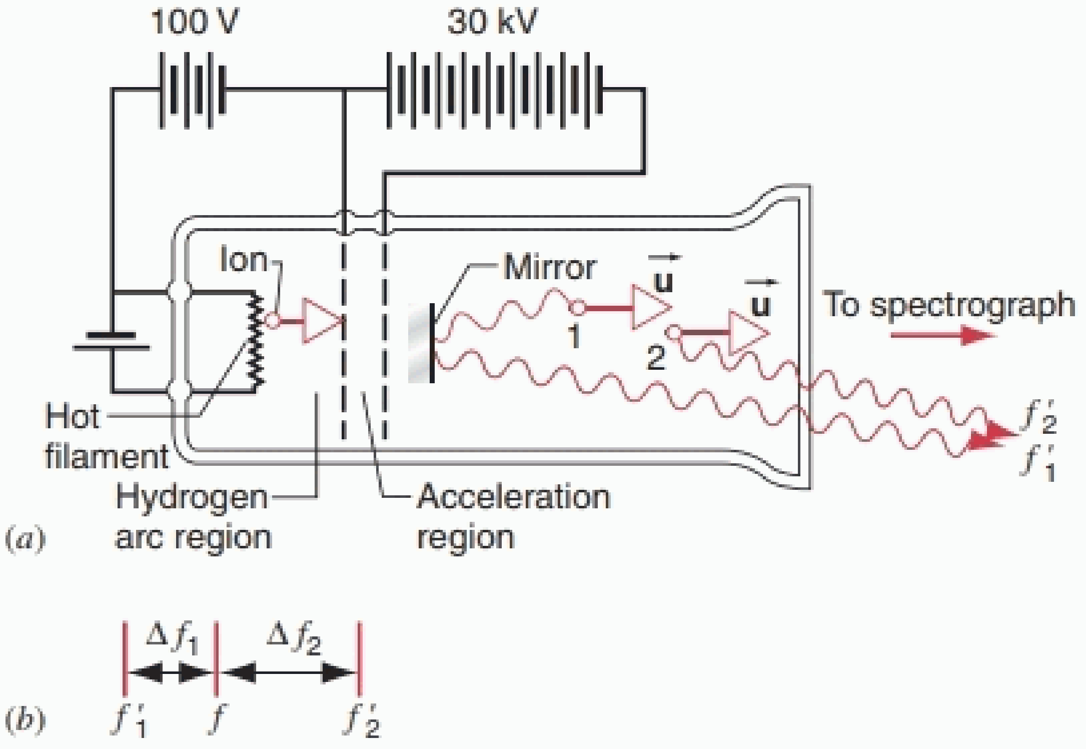


FIGURE 39-27. Apparatus used in the Ives–Stilwell experiment.

TABLE 39-4 The Ives–Stilwell Experiment

Speed u	$\Delta f/f\ 10^{-5}$		
(10^6 m/s)	*Classical*	*Relativistic*	*Experiment*
0.865	1.67	0.835	0.762
1.01	2.26	1.13	1.1
1.15	2.90	1.45	1.42
1.33	3.94	1.97	1.9

rest; such a line appears as f in Fig. 39-27*b*. A fundamental measured quantity in this experiment is $\Delta f/f$, defined from

$$\frac{\Delta f}{f} = \frac{\Delta f_2 - \Delta f_1}{f}, \tag{39-22}$$

(see Fig. 39-27*b*). It measures the extent to which the frequency of the light from resting atoms fails to lie halfway between the frequencies f_1' and f_2'. Table 39-4 shows that the measured results agree with the formula predicted by the theory of relativity (Eq. 39-20) and not with the classical formula borrowed from the theory of sound propagation in a material medium (Eq. 39-18).

Ives and Stilwell did not present their experimental results as evidence for the support of Einstein's theory of relativity, but rather gave them an alternative theoretical explanation. Modern observers, looking not only at their excellent experiment but at the whole range of experimental evidence, now give the Ives–Stilwell experiment the interpretation we have described, as a test of the relativistic Doppler effect.

SAMPLE PROBLEM 39-8. A distant quasar is moving away from Earth at a speed u. An astronomer is searching for a certain spectral line in the light from the quasar. That line, emitted by atomic hydrogen, is observed using hydrogen discharge tubes on Earth to have a wavelength of $\lambda_0 = 121.6$ nm. The astronomer finds the hydrogen spectral line emitted by the quasar at a wavelength of $\lambda = 460.9$ nm. Assuming the quasar to be moving radially away from Earth, what is its speed relative to the Earth?

Solution We use Eq. 39-20, which we rewrite in terms of wavelength:

$$\frac{c}{\lambda} = \frac{c}{\lambda_0}\frac{1 - u/c}{\sqrt{1 - u^2/c^2}}$$

or

$$\lambda = \lambda_0\sqrt{\frac{1 + u/c}{1 - u/c}}. \tag{39-23}$$

Solving for the speed, we find

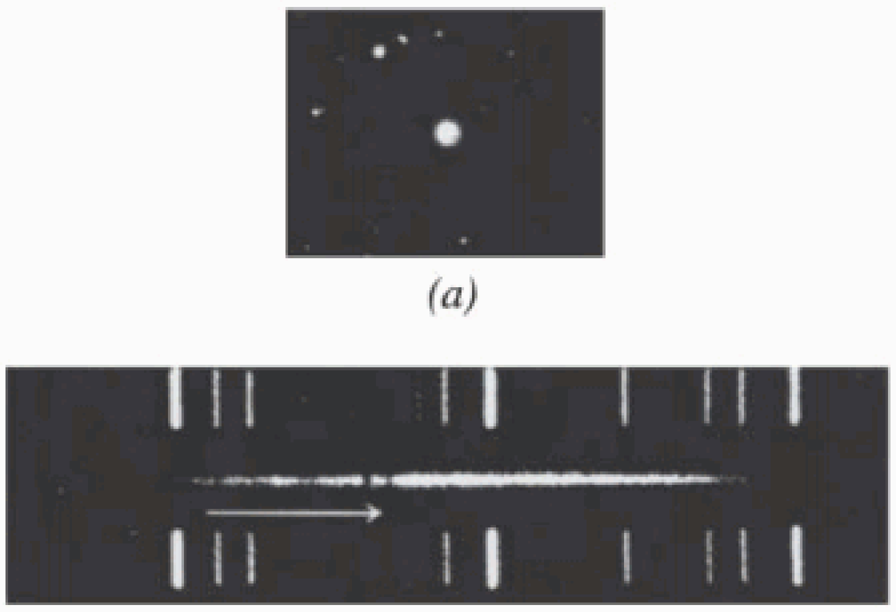
$$\frac{u}{c} = \frac{(\lambda/\lambda_0)^2 - 1}{(\lambda/\lambda_0)^2 + 1}. \tag{39-24}$$

With $\lambda/\lambda_0 = 460.9\ \text{nm}/121.6\ \text{nm} = 3.79$, we obtain

$$\frac{u}{c} = \frac{(3.79)^2 - 1}{(3.79)^2 + 1} = 0.87.$$

The quasar is moving away from Earth at 87% of the speed of light. This calculation determines only the radial or line-of-sight component of the relative velocity.

The Doppler effect causes the wavelengths of light from objects receding from Earth to be lengthened or shifted toward the red (long-wavelength) end of the visible spectrum. Hence it is known as the *red shift.* Figure 39-28 shows an example of a red-shifted spectrum, from which it is possible to determine the speed of the galaxy relative to the Earth. Evidence from many such observations shows that all distant objects are moving away from us, and that there is a direct (linear) relationship between the speed of the object and its distance from Earth: the more distant the object, the faster it moves away from us. This linear behavior, deduced from measurements of the red shift, is the primary evidence for the expansion of the universe.


(*a*)

(*b*)

FIGURE 39-28. (*a*) A galaxy in the constellation Corona Borealis. (*b*) The central streak shows the wavelength spectrum of the light emitted by this galaxy. The two dark vertical bands show absorption lines associated with calcium, which is present in the galaxy. The line spectra above and below are recorded from a laboratory source to provide a wavelength calibration. The horizontal arrow shows how far the calcium lines are displaced from where they would be expected to appear if they were emitted by a source at rest in the laboratory. From this Doppler shift, the recessional speed of the galaxy is deduced to be about 21,000 km/s.

Derivation of the Relativistic Doppler Effect

Let us now look at the origin of Eq. 39-20 for the relativistic Doppler effect and why it differs from Eqs. 39-18 and 39-19. Figure 39-29*a* shows an observer S' who carries a source of electromagnetic waves. This source emits a series of N wavefronts. Observer S is in motion relative to S' with velocity $\vec{\mathbf{u}}$. At the instant of time shown in Fig. 39-29*a*, the first wavefront has just arrived at the present position of S. According to S', the distance traveled by that wavefront is $c\,\Delta t_0$. We use the subscript 0 to emphasize that the interval Δt_0 is a *proper time*; as defined in Chapter 20, a proper time interval is one in which the events marking the beginning and end of the interval occur at the same location. The interval Δt_0 is measured by S' as the time between the emission of the first wavefront and the Nth wavefront, which occur at the same location in the reference frame of S'.

According to S', the wavelength λ' is the total distance occupied by the N wavefronts divided by the number of wavefronts, or

$$\lambda' = \frac{c\,\Delta t_0}{N}. \tag{39-25}$$

The frequency and wavelength are related by $c = \lambda' f_0$, so

$$f_0 = \frac{c}{\lambda'} = \frac{N}{\Delta t_0}. \tag{39-26}$$

Now let us look at the same situation from the viewpoint of observer S, relative to whom S' is in motion with velocity $-\vec{\mathbf{u}}$, as shown in Fig. 39-29*b*. According to S, the first wavefront travels from the original location of S' to the (fixed) location of S in a time Δt, and it travels a distance $c\,\Delta t$. In that same time interval, S' moves a distance $u\,\Delta t$. (Note that Δt is *not* a proper time interval for S, who observes the first wavefront to be emitted when the source is at one location and the final wavefront to originate with the source at a different location.)

According to S, the N wavefronts occupy a distance of $u\,\Delta t + c\,\Delta t$, so the wavelength λ is

$$\lambda = \frac{u\,\Delta t + c\,\Delta t}{N}. \tag{39-27}$$

With $f = c/\lambda$ and $N = f_0\,\Delta t_0$ from Eq. 39-26, we have

$$f = \frac{c}{\lambda} = \frac{cN}{u\,\Delta t + c\,\Delta t} = f_0 \frac{\Delta t_0}{\Delta t}\,\frac{1}{1 + u/c}. \tag{39-28}$$

The relationship between the proper time interval Δt_0 and the (longer) time interval Δt is given by the time dilation formula, Eq. 20-3: $\Delta t = \Delta t_0/\sqrt{1 - u^2/c^2}$, so

$$f = f_0 \frac{\sqrt{1 - u^2/c^2}}{1 + u/c} = f_0 \frac{1 - u/c}{\sqrt{1 - u^2/c^2}}, \tag{39-29}$$

which is identical to Eq. 39-20.

Why does this derivation for light waves give a result that differs from the ones for classical (sound) waves? In Eqs. 39-26 and 39-28, we have used $c = \lambda' f_0$ and $c = \lambda f$, indicating that observers S' and S measure the same speed for the light waves. This is required by the second principle of relativity, which asserts that the speed of light in empty space has the same value for all observers, no matter what their state of relative motion. For sound waves, however, the wave speed relative to the medium carrying the wave varies for observers moving with different speeds.

The Transverse Doppler Effect

We have derived Eq. 39-29 for the special case in which the relative motion of S and S' is along the line connecting them. In general, the velocity $\vec{\mathbf{u}}$ might be at some angle θ relative to this line. In the first term of Eq. 39-29, the numerator comes from the time dilation factor and depends only on the magnitude of $\vec{\mathbf{u}}$ and not on the relative direction of the motion. The denominator, however, involves the component of $\vec{\mathbf{u}}$ along the line connecting S and S', so that the distance $u\,\Delta t$ in Fig. 39-29*b* and Eq. 39-28 would be $(u\cos\theta)\,\Delta t$. As a result, Eq. 39-29 becomes

$$f = f_0 \frac{\sqrt{1 - u^2/c^2}}{1 + (u/c)\cos\theta}. \tag{39-30}$$

For the special case in which $\theta = 90°$, Eq. 39-30 is

$$f = f_0\sqrt{1 - u^2/c^2}. \tag{39-31}$$

Equation 39-31 describes the *transverse Doppler effect,* which applies to the case in which the source or the observer is moving transverse or perpendicular to the line connecting them. The classical Doppler effect for sound waves has no transverse term; there is no change in frequency if the source of sound or the observer moves transversely. The transverse Doppler effect is a purely relativistic effect with no classical counterpart. A beautiful and sensitive confirmation of the transverse Doppler effect for gamma rays was done in 1963 by Walter Kundig, using a source of gamma rays at the center of a centrifuge-like rotor and a detector on the rim. A summary of his results is presented in Fig. 39-30.

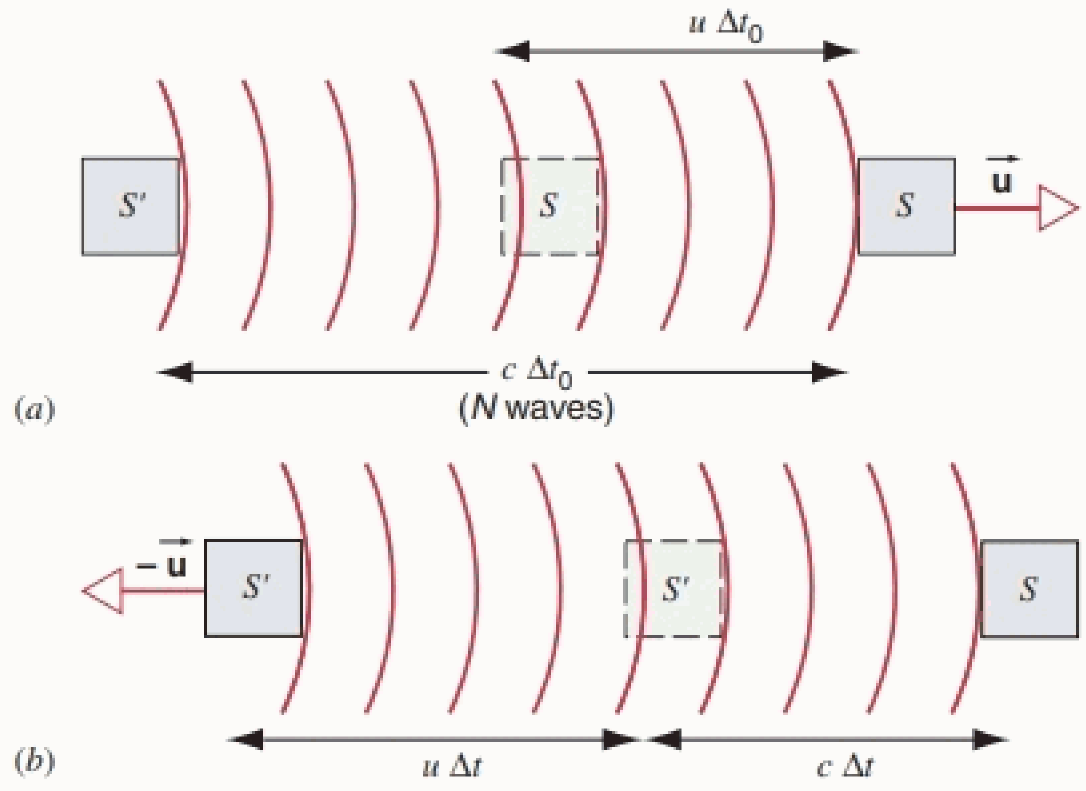


FIGURE 39-29. (*a*) A source of light waves in reference frame S' emits N wavefronts. The figure shows the instant the first wavefront arrives at S, which is in motion with velocity $\vec{\mathbf{u}}$. (*b*) The same situation from the frame of reference of S.

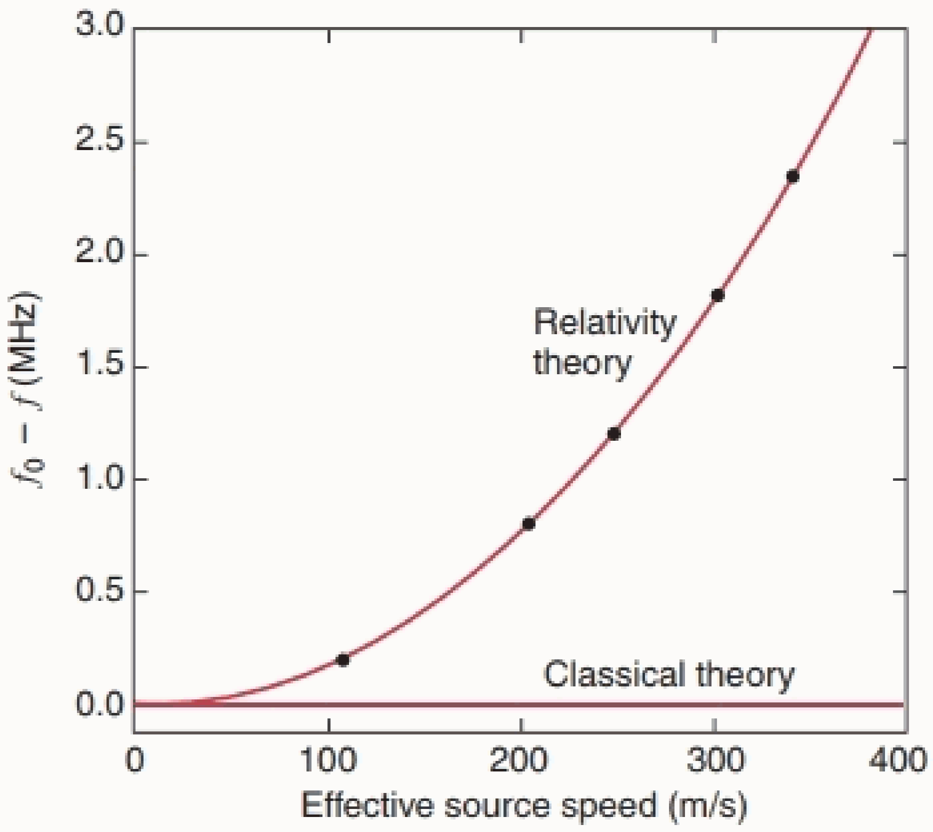


FIGURE 39-30. The results of Kundig's experiment (data points indicated by black dots) for the transverse Doppler effect agree with relativity theory (Eq. 39-16) and disagree with classical theory, which predicts no effect.

SAMPLE PROBLEM 39-9. An Earth satellite is orbiting from west to east at an altitude of $h = 153$ km in a circular orbit above the equator (see Fig. 39-31). A tracking ship is located at the equator on the Prime Meridian at 0° longitude (just off the west coast of Africa). The satellite emits radio waves at a frequency of 122.450 MHz. To what frequency should the ship tune its receiver when the satellite is (*a*) directly overhead, (*b*) at 10° longitude west of the ship, and (*c*) at 10° longitude east of the ship?

Solution (*a*) We let the S' frame be moving with the satellite at the instant it is overhead; the S frame is that of the ship below. The frequency f_0 observed in the S' frame (the satellite) is 122.450 MHz. The relative speed u between the frames is determined by the orbital speed of the satellite at an altitude h or at a radius $R = R_E + h$, where R_E is the radius of the Earth. The gravitational acceleration at a radius R is MG/R^2, which must supply the centripetal acceleration u^2/R necessary for a circular orbit. Thus $MG/R^2 = u^2/R$, or

$$u = \sqrt{\frac{MG}{R}} = \sqrt{\frac{MG}{R_E + h}}$$
$$= \sqrt{\frac{(5.98 \times 10^{24}\text{ kg})(6.67 \times 10^{-11}\text{ N}\cdot\text{m}^2/\text{kg}^2)}{6370\text{ km} + 153\text{ km}}}$$
$$= 7.82 \times 10^3\text{ m/s} = 2.61 \times 10^{-5}c.$$

When the satellite is directly overhead, the Doppler shift is obtained from Eq. 39-30 with $\theta = 90°$:

$$f = f_0\sqrt{1 - u^2/c^2}.$$

With $u/c = 2.61 \times 10^{-5}$, we have $u^2/c^2 = 6.8 \times 10^{-10}$. The quantity under the radical differs from 1 by only a few parts in 10^{10}, so that, to the desired precision,

$$f \approx f_0 = 122.450\text{ MHz}.$$

(*b*) When the satellite is not overhead, it is necessary to calculate the angle θ between the velocity of the satellite and the direct line to the tracking ship (see Fig. 39-31). We draw line BD perpendicular to AC $(= R_E + h)$. Then, because angle $ABD = \theta$, we have $\cot\theta = BD/AD$, where $BD = R_E\sin\phi$ and $AD = AC - CD = R_E + h - R_E\cos\phi$, so

$$\cot\theta = \frac{R_E\sin\phi}{h + R_E(1 - \cos\phi)}$$
$$= \frac{(6370\text{ km})(\sin 10°)}{153\text{ km} + (6370\text{ km})(1 - \cos 10°)} = 4.428,$$

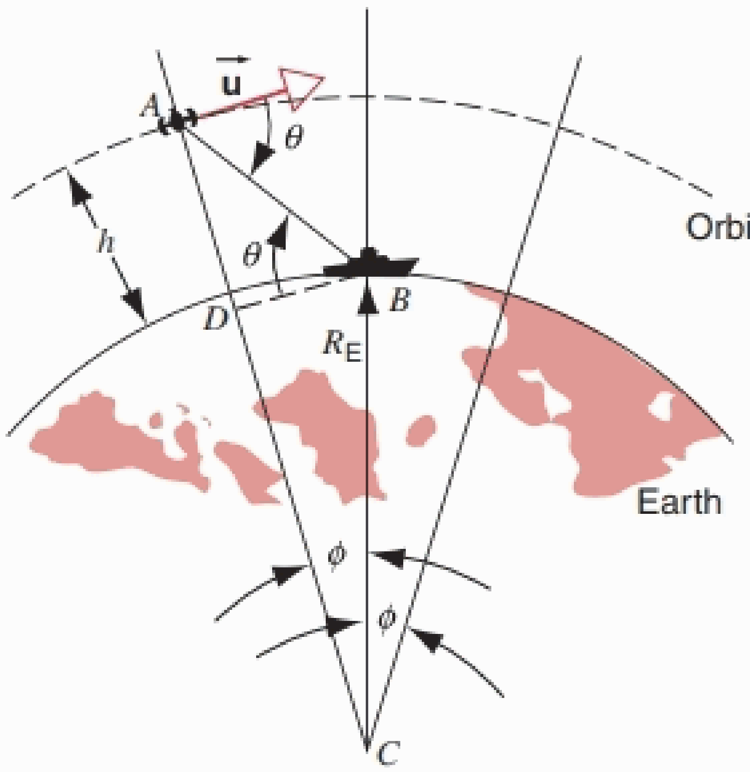


FIGURE 39-31. Sample Problem 39-9. A satellite is in a circular orbit at an altitude h above the surface of the Earth. A ship on the surface observes signals beamed by the satellite.

or

$$\theta = \cot^{-1} 4.428 = 12.7°.$$

We can now use Eq. 39-30 to find the frequency, using our result from part (*a*) that the square-root factor is nearly 1 and substituting $-u$ for u because the satellite is approaching the ship:

$$f \approx \frac{f_0}{1 - (u/c)\cos\theta} = \frac{122.450\text{ MHz}}{1 - (2.61 \times 10^{-5})(\cos 12.7°)}.$$
$$= 122.453\text{ MHz}.$$

(*c*) After the satellite passes over the ship, its motion becomes away from the observer, and Eq. 39-15 gives

$$f \approx \frac{f_0}{1 + (u/c)\cos\theta} = \frac{122.450\text{ MHz}}{1 + (2.61 \times 10^{-5})(\cos 12.7°)}$$
$$= 122.447\text{ MHz}.$$

We see that the frequency detected on Earth varies from 122.453 MHz (satellite approaching) to 122.450 MHz (satellite overhead) to 122.447 MHz (satellite receding). A measurement of the Doppler-shifted frequency is thus sufficient to locate the satellite.

MULTIPLE CHOICE

39-1 The Electromagnetic Spectrum

39-2 Visible Light

1. Which color of light is *least* absorbed by plants?
   (A) Red (B) Yellow (C) Green (D) Violet

39-3 The Speed of Light

2. The speed of light in water is less than the speed of light in air. When a ray of red light, $\lambda = 650$ nm, moves from air into water, what changes?
   (A) The frequency (B) The wavelength
   (C) The color (D) Both A and B
   (E) A, B, and C

3. In which of the following types of substances could light have a speed greater than c?
   (A) Dielectric (B) Diamagnetic
   (C) Paramagnetic (D) Ferromagnetic
   (E) None of these

4. According to the discussion of Section 39-3, would you expect red light to move faster or slower through a substance than blue light?

(A) The red light would move faster.
(B) The blue light would move faster.
(C) Both move at the same speed, since color is a psychological label.
(D) The answer depends on the substance.

39-4 Reflection and Refraction of Light Waves

5. A ray of light strikes a corner reflector as shown in Fig. 39-13. As the angle of incidence θ is increased, the angle between the final reflected ray and the incident ray

(A) increases. (B) decreases.
(C) remains the same.
(D) The question cannot be answered without solving each incident angle individually.

6. A ray of light strikes one of two mirrors that meet at a 60° angle as shown in Fig. 39-32. As the angle of incidence θ is increased, the angle between the final reflected ray and the incident ray

(A) increases. (B) decreases. (C) remains the same.
(D) The question cannot be answered without solving each incident angle individually.

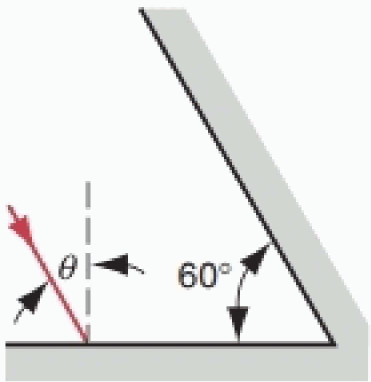


FIGURE 39-32. Multiple-choice question 6.

7. A ray of light shines into two mirrors that meet at an angle α (measured in radians). What is a reasonable estimate of the maximum number of times the light will bounce off of the mirrors before it comes back out?

(A) $2\pi/\alpha$ (B) π/α (C) $\sqrt{\pi/\alpha}$ (D) $\sqrt{\alpha/\pi}$

8. A ray of light is incident on a boundary between two substances as shown in Fig. 39-33. The substances have different indices of refraction. Which outgoing ray could *not* be produced from the incident ray?

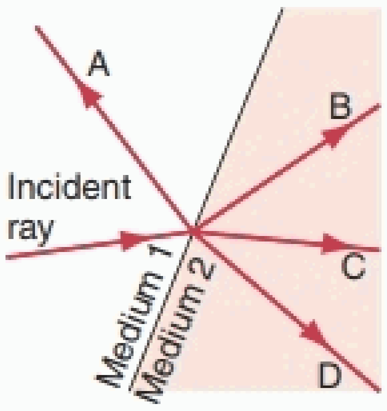


FIGURE 39-33. Multiple-choice question 8.

9. A ray of light is incident on a boundary between two substances as shown in Fig. 39-34. The substances have different indices of refraction. Which incident ray could *not* have produced the outgoing ray?

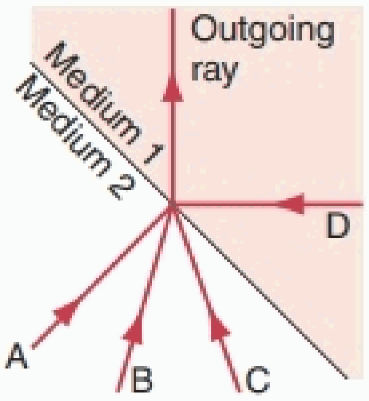


FIGURE 39-34. Multiple-choice question 9.

10. The angle of incidence for a light ray is $\theta_i = 20°$, and the angle of refraction for the refracted ray is $\theta_r > 20°$.

(*a*) If the angle of incidence is decreased by 10°, then the angle of refraction

(A) will also decrease by 10°.
(B) will decrease by more than 10°.
(C) will decrease by less than 10°.
(D) The question cannot be answered without more information.

(*b*) If instead the angle of incidence is changed so that the angle of refraction doubles, then the angle of incidence

(A) must also have doubled.
(B) will more than double.
(C) will less than double.
(D) The question cannot be answered without more information.

39-5 Total Internal Reflection

11. Light traveling through three transparent substances follows the path shown in Fig. 39-35. Arrange the indices of refraction in order from smallest to largest. Note that total internal reflection *does* occur on the bottom surface of medium 2.

(A) $n_1 < n_2 < n_3$ (B) $n_2 < n_1 < n_3$
(C) $n_1 < n_3 < n_2$ (D) $n_3 < n_1 < n_2$

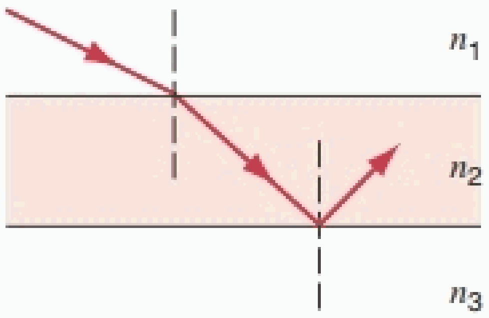


FIGURE 39-35. Multiple-choice question 11.

12. Light is incident on a glass block as shown in Fig. 39-36. θ_1 is increased slightly. What happens to θ_2?
(A) θ_2 also increases slightly.
(B) θ_2 is unchanged.
(C) θ_2 decreases slightly.
(D) θ_2 changes abruptly, since the ray experiences total internal reflection.

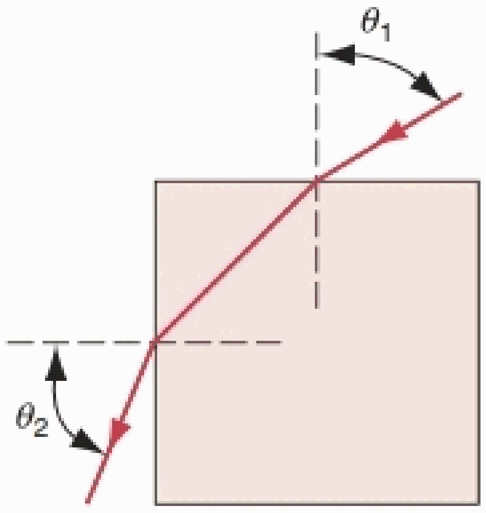


FIGURE 39-36. Multiple-choice question 12.

13. A lightbulb is immersed in water. Light travels out in all directions from the bulb, but only some of the light escapes the water's surface. (Compare with Fig. 39-21.) What happens to the fraction f of light that escapes the water's surface as the bulb is moved deeper into the water?
(A) f increases. (B) f decreases.
(C) f remains the same.
(D) The answer depends on the index of refraction of water.

14. A ray of light is observed to exit from a boundary between two substances as shown in Fig. 39-37. The substances have different indices of refraction. Which incident ray could *not* have produced the outgoing ray?

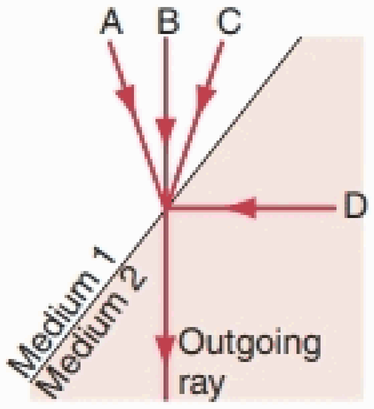


FIGURE 39-37. Multiple-choice question 14.

39-6 The Doppler Effect for Light

15. Traveling at a speed v away from Earth you notice that a radio signal from Earth is Doppler-shifted from 96 MHz to 48 MHz. You turn around and head back to Earth at the same speed. What will be the Doppler-shifted frequency of the radio signal now?
(A) Greater than 192 MHz
(B) Equal to 192 MHz
(C) Less than 192 MHz

QUESTIONS

1. Electromagnetic waves reach us from far into space. From the information they carry, can we tell what the universe is like at the present moment? At any selected time in the past?

2. If you were asked on an examination what fraction of the electromagnetic spectrum lies in the visible range, what would you reply?

3. List several ways in which radio waves differ from visible light waves. In what ways are they the same?

4. How would you characterize electromagnetic radiation that has a frequency of 10 kHz? 10^{20} Hz? A wavelength of 500 nm? 10 km? 0.50 nm?

5. What determines the desirable length and orientation of the "rabbit ears" antenna on a TV set?

6. How does a microwave oven cook food? You can boil water in a plastic bag in such an oven. How can this happen?

7. How might an eye-sensitivity curve like that of Fig. 39-6 be measured?

8. Why are danger signals in red, when the eye is most sensitive to yellow-green?

9. In connection with Fig. 39-6, (*a*) do you think it possible that the wavelength of maximum sensitivity could vary if the intensity of the light is changed? (*b*) What might the curve of Fig. 39-6 look like for a group of color-blind people who could not, for example, distinguish red from green?

10. Suppose that human eyes were insensitive to visible light but were very sensitive to infrared light. What environmental changes would be needed if you were to (*a*) walk down a long corridor and (*b*) drive a car? Would the phenomenon of color exist? How might traffic lights have to modified?

11. What feature of light corresponds to loudness in sound?

12. How could Galileo have tested experimentally that reaction times were an overwhelming source of error in his attempt to measure the speed of light, described in Section 39-3?

13. Can you think of any "everyday" observation (that is, without experimental apparatus) to show that the speed of light is not infinite? Think of lightning flashes, possible discrepancies between the predicted time of sunrise and the observed time, radio communications between Earth and astronauts in orbiting spaceships, and so on.

14. Comment on this statement: Because of the way the meter is defined, it is no longer possible to measure the speed of light.

15. It has been suggested that the speed of light may change slightly in value as time goes on. Can you find any evidence for this in Table 39-1?

16. In a vacuum, does the speed of light depend on (*a*) the wavelength, (*b*) the frequency, (*c*) the intensity, (*d*) the speed of the source, or (*e*) the speed of the observer?

17. Describe what your immediate environment would be like if all objects were totally absorbing of light. Sitting in a chair in a room, could you see anything? If a cat entered the room could you see it?

18. Can you think of a simple test or observation to prove that the law of reflection is the same for all wavelengths, under conditions in which geometrical optics prevail?

19. A street light, viewed by reflection across a body of water in which there are ripples, appears very elongated in the line of vision but not sideways. Explain.

20. Shortwave broadcasts from Europe are heard in the United States even though the path is not a straight line. Explain how.

21. The travel time of signals from satellites to receiving stations on Earth varies with the frequency of the signal. Why?

22. By what percentage does the speed of blue light in fused quartz differ from that of red light?

23. How can one determine the indices of refraction of the media in Table 39-3 relative to water, given the data in that table?

24. Would you expect sound waves to obey the laws of reflection and refraction obeyed by light waves? Discuss the propagation of spherical and cylindrical waves using Huygens' principle. Does Huygens' principle apply to sound waves in air?

25. If Huygens' principle predicts the laws of reflection and refraction, why is it necessary or desirable to view light as an electromagnetic wave, with all its attendant complexity?

26. A light beam is broadened on entering water. Explain.

27. Is it correct to say that there is no interaction between visible light and a transparent medium through which it passes?

28. How does atmospheric refraction affect the apparent time of sunset?

29. Stars twinkle but planets do not. Why?

30. Explain why the far end of a pool filled to a uniform depth appears shallower than the end near the observer at its edge.

31. It is a bright, sunny day and you want to create a rainbow in your backyard using a garden hose. Exactly how do you go about it? Incidentally, why can you not walk under, or go to the end of, a rainbow?

32. Is it possible, by using one or more prisms, to recombine into white light the color spectrum formed when white light passes through a single prism? If yes, explain how.

33. Describe and explain what a fish sees as it looks in various directions above its horizon.

34. Why does a diamond "sparkle" more than a glass imitation cut to the same shape?

35. Is it plausible that the wavelength of light should change in passing from air into glass but that its frequency should not? Explain.

36. In reflection and refraction, why do the reflected and refracted rays lie in the plane defined by the incident ray and the normal to the surface? Can you think of any exceptions?

37. Design a periscope, taking advantage of total internal reflection. What are the advantages compared with silvered mirrors?

38. What characteristics must a material have in order to serve as an efficient "light pipe"?

39. A certain toothbrush has a red plastic handle into which rows of nylon bristles are set. The tops of the bristles (but not their sides) appear red. Explain.

40. Why are optical fibers more effective carriers of information than, say, microwaves or cables? Think of the frequencies involved.

41. Is the Doppler effect simply a time-dilation effect and nothing more, or is there something else to it?

42. One member of a binary star system emits visible light. Show on a simple graph how the Doppler frequency shift on Earth varies with time.

43. Can you a galaxy be so distant that its recession speed equals c? If so, how can we see the galaxy? That is, will its light ever reach us?

44. Gamma rays are electromagnetic radiation emitted from radioactive nuclei. In free space, do they travel with the same speed as visible light? Does their speed depend on the speed of the nucleus that emits them?

45. Perhaps the simplest astronomical observation that you can make is this: When the Sun sets, the sky becomes dark. This is true and seems obvious, but an argument can be made that it should not be so. Consider: "Assuming an infinite universe, uniformly populated by stars more or less like our Sun, we can say that a straight line projected from the observer in any direction will eventually hit a star. The distances R of most of these stars will be very great indeed, so that the stars illuminate the observer only very weakly, the illumination varying as $1/R^2$. On the other hand, the number of distant stars located within a spherical shell whose radii are R and $R + dR$ increases as R^2 (assuming that dR is constant). Can you prove this last statement? These two effects seem to cancel precisely. Thus the night sky should be virtually infinitely bright, the observer being illuminated by an infinity of suns." Can you see any flaw in this argument (usually called Olber's paradox)? Think of the finite speed of light, the large scale of the universe, the expanding universe and the associated red shift, the finite lifetime of stars, and so on. (See "The Dark Sky Paradox," by E. R. Harrison, *American Journal of Physics,* February 1977, p. 119, for an excellent historical review and a lucid explanation.)

EXERCISES

39-1 The Electromagnetic Spectrum

1. Show that the frequency and wavelength markings on Fig. 39-1 obey the relation $f\lambda = c$.

2. Project Seafarer was an ambitious program to construct an enormous antenna, buried underground on a site about 4000 square miles in area. Its purpose was to transmit signals to submarines while they were deeply submerged. If the effective wavelength was 1.0×10^4 Earth radii, what would be (*a*) the frequency and (*b*) the period of the radiations emitted? Ordinarily, electromagnetic radiations do not penetrate very far into conductors such as seawater. Can you think of any reason why such ELF (extremely low frequency) radiations should penetrate more effectively? Think of the limiting case of zero frequency. (Why not transmit signals at zero frequency?)

3. (*a*) The wavelength of the most energetic x rays produced when electrons accelerated to 30 GeV in the Stanford Linear Accelerator slam into a solid target is 0.067 fm. What is the frequency of these x rays? (*b*) A VLF (very low frequency) radio wave has a frequency of only 30 Hz. What is its wavelength?

4. The radiation from a certain HeNe laser, although centered on 632.8 nm, has a finite "linewidth" of 0.010 nm. Calculate the linewidth in frequency units.

39-2 Visible Light

5. (*a*) At what wavelengths does the eye of a standard observer have half its maximum sensitivity? (*b*) What are the wavelength, the frequency, and the period of the light for which the eye is the most sensitive?

6. How many complete vibrations are contained in the wave train of light of wavelength 520 nm emitted by a laser for a time of 430 ps?

39-3 The Speed of Light

7. (*a*) Suppose that we were able to establish radio communication with the hypothetical inhabitants of a hypothetical planet orbiting our nearest star, α-Centauri, which is 4.34 light-years from us. How long would it take to receive a reply to a message? (*b*) Repeat for the Great Nebula in Andromeda, one of our closest extragalactic neighbors but 2.2×10^6 light-years distant. What do these considerations lead you to conclude about the nature of our possible communication with extragalactic civilizations?

8. (*a*) How long does it take a radio signal to travel 150 km from a transmitter to a receiving antenna? (*b*) We see a full Moon by reflected sunlight. How much earlier did the light that enters our eye leave the Sun? (*c*) What is the round-trip travel time for light between Earth and a spaceship orbiting Saturn, 1.3×10^9 km distant? (*d*) The Crab Nebula, which is about 6500 light-years distant, is thought to be the result of a supernova explosion recorded by Chinese astronomers in A.D. 1054. In approximately what year did the explosion actually occur?

9. The uncertainty of the distance to the Moon, as measured by the reflection of laser light from reflectors placed on the Moon by Apollo 11 astronauts, is about 2 cm. This uncertainty is associated with the measurement of the elapsed time; what uncertainty in this time is implied?

39-4 Reflection and Refraction of Light Waves

10. A corner reflector, much used in optical, microwave, and other applications, consists of three plane mirrors fastened together as the corner of a cube. It has the property that an incident ray is returned, after three reflections, with its direction exactly reversed. Prove this result.

11. Using Fermat's principle, prove that the reflected ray, the incident ray, and the normal lie in one plane.

12. One end of a stick is dragged through water at a speed v that is greater than the speed u of water waves. Applying Huygens' construction to the water waves, show that a conical wavefront is set up and that its half-angle α is given by

$$\sin \alpha = u/v.$$

This is familiar as the bow wave of a ship or the shock wave caused by an object moving through air with a speed exceeding that of sound, as in Fig. 19-16.

13. In Fig. 39-38, find the angles (*a*) θ_1 and (*b*) θ_2.

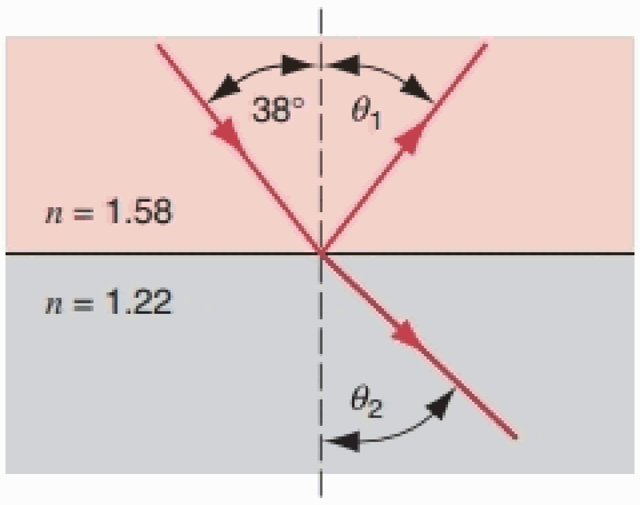


FIGURE 39-38. Exercise 13.

14. Light in vacuum is incident on the surface of a glass slab. In the vacuum the beam makes an angle of 32.5° with the normal to the surface, while in the glass it makes an angle of 21.0° with the normal. Find the index of refraction of the glass.

15. The speed of yellow sodium light in a certain liquid is measured to be 1.92×10^8 m/s. Find the index of refraction of this liquid with respect to air, for sodium light.

16. Find the speed in fused quartz of light of wavelength 550 nm. (See Fig. 39-11.)

17. When an electron moves through a medium at a speed exceeding the speed of light in that medium, it radiates electromagnetic waves (the Cerenkov effect). What minimum speed must an electron have in a liquid of index of refraction 1.54 in order to radiate?

18. A laser beam travels along the axis of a straight section of pipeline 1.61 km long. The pipe normally contains air at standard temperature and pressure, but it may also be evacuated. In which case would the travel time for the beam be greater and by how much?

19. When the rectangular metal tank in Fig. 39-39 is filled to the top with an unknown liquid, an observer with eyes level with the top of the tank can just see the corner *E*. Find the index of refraction of the liquid.

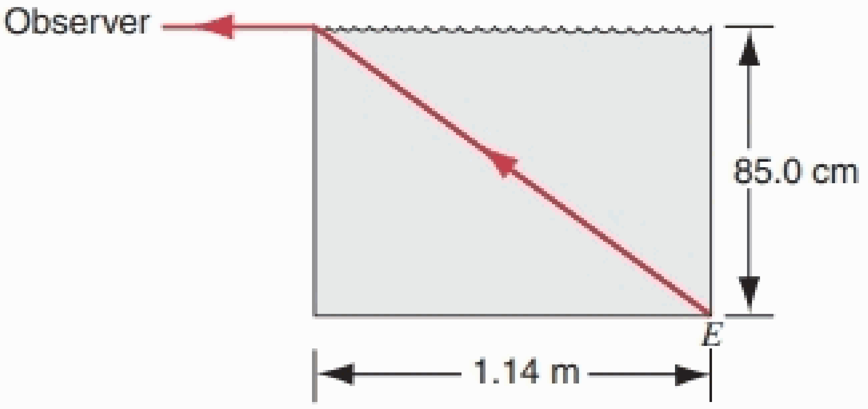


FIGURE 39-39. Exercise 19.

20. A ray of light goes through an equilateral prism in the position of minimum deviation. The total deviation is 37°. What is the index of refraction of the prism? See Sample Problem 39-4.

21. In Fig. 39-14 (Sample Problem 39-4) show by graphical ray tracing, using a protractor, that if θ for the incident ray is either increased or decreased, the deviation angle ψ is increased.

22. Light from a laser enters a glass block at A and emerges at B; see Fig. 39-40. The glass block has a length $L = 54.7$ cm and an index of refraction $n = 1.63$. The angle of incidence is $\theta = 24.0°$. Find the time needed for light to pass through the block.

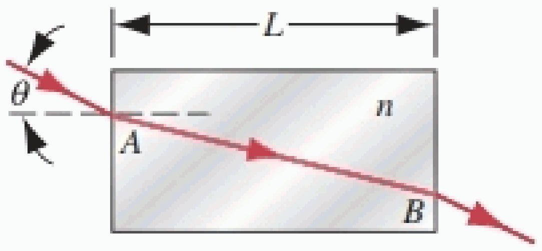


FIGURE 39-40. Exercise 22.

23. A diver beneath the surface of water in a lake looks up at 27° from the vertical to see a life ring floating on the surface. Through the center of the ring can be seen the top of a smokestack known to be 98 m high. How far is the base of the smokestack from the life ring?

24. A bottom-weighted 200-cm-long vertical pole extends from the bottom of a swimming pool to a point 64 cm above the water. Sunlight is incident at 55° above the horizon. Find the length of the shadow of the pole on the level bottom of the pool.

25. A coin lies at the bottom of a pool with depth d and index of refraction n, as shown in Fig. 39-41. Show that light rays that are close to the normal appear to come from a point $d_{\text{app}} = d/n$ below the surface. This distance is the apparent depth of the pool.

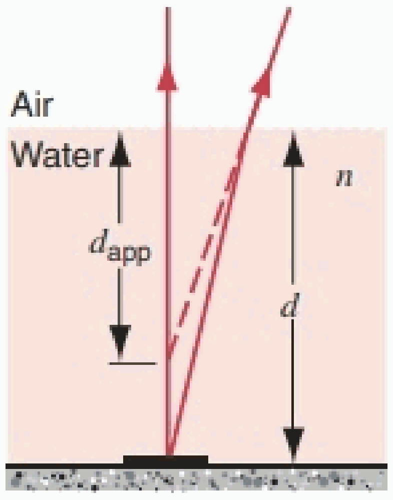


FIGURE 39-41. Exercise 25.

26. A glass prism with an apex angle of 60° has $n = 1.60$. (a) What is the smallest angle of incidence for which a ray can enter one face of the prism and emerge from the other? (b) What angle of incidence would be required for the ray to pass through the prism symmetrically? See Sample Problem 39-4.

27. A layer of water ($n = 1.33$) 20 mm thick floats on a layer of carbon tetrachloride ($n = 1.46$) 41 mm thick. How far below the water surface, viewed at near normal incidence, does the bottom of the tank seem to be?

28. The apparent depth of a pool depends on the angle of viewing. Suppose that you place a coin at the bottom of a swimming pool filled with water ($n = 1.33$) to a depth of 2.16 m. Find the apparent depth of the coin below the surface when viewed (a) at near normal incidence and (b) by rays that leave the coin making an angle of 35.0° with the normal to the bottom of the pool. See Exercise 25.

29. Light of wavelength 612 nm in a vacuum travels 1.57 pm in a medium of index of refraction 1.51. Find (a) the wavelength in the medium, (b) the optical path length, and (c) the phase difference after moving that distance, with respect to light traveling the same distance in a vacuum.

30. The wavelength of yellow sodium light in air is 589 nm. (a) What is its frequency? (b) What is its wavelength in glass whose index of refraction is 1.53? (c) From the results of (a) and (b), find its speed in this glass.

31. Prove that the optical path lengths for reflection and refraction in Figs. 39-17 and 39-19 are each a minimum when compared with other nearby paths connecting the same two points. (Hint: Examine the quantity d^2L/dx^2.)

39-5 Total Internal Reflection

32. A ray of light is incident normally on the face ab of a glass prism ($n = 1.52$), as shown in Fig. 39-42. (a) Assuming that the prism is immersed in air, find the largest value for the angle ϕ so that the ray is totally reflected at face ac. (b) Find ϕ if the prism is immersed in water.

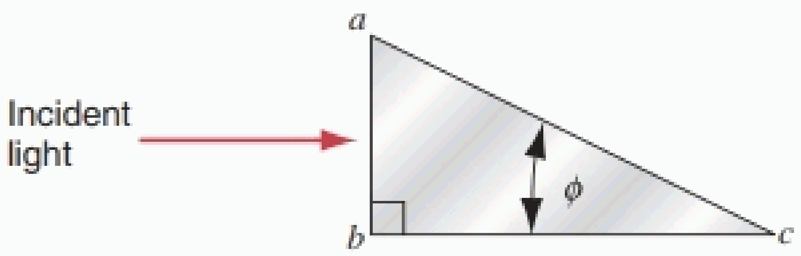


FIGURE 39-42. Exercise 32.

33. Two materials, A and B, have indices of refraction of 1.667 and 1.586, respectively. (a) Find the critical angle for total internal reflection at an interface between the two materials. (b) In which direction must an incident ray be propagating if it is to be totally reflected?

34. A fish is 1.8 m below the surface of a smooth lake. At what angle above the horizontal must it look to see the light from a small camp fire burning at the water's edge 92 m away?

35. A point source of light is 82.0 cm below the surface of a body of water. Find the diameter of the largest circle at the surface through which light can emerge from the water.

36. A light ray falls on a square glass slab as in Fig. 39-43. What must be the minimum index of refraction of the glass if total internal reflection occurs at the vertical face?

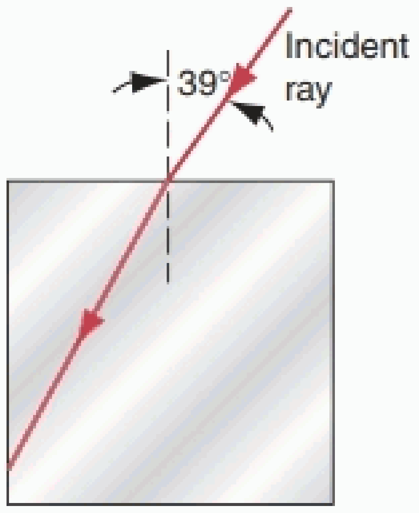


FIGURE 39-43. Exercise 36.

37. A plane wave of white light traveling in fused quartz strikes a plane surface of the quartz, making an angle of incidence θ. Is it possible for the internally reflected beam to appear (*a*) bluish or (*b*) reddish? (*c*) Roughly what value of θ must be used? (Hint: White light will appear bluish if wavelengths corresponding to red are removed.)

38. A glass cube has a small spot at its center. (*a*) What parts of the cube face must be covered to prevent the spot from being seen, no matter what the direction of viewing? (*b*) What fraction of the cube surface must be so covered? Assume a cube edge of 12.6 mm and an index of refraction of 1.52. (Neglect the subsequent behavior of an internally reflected ray.)

39-6 The Doppler Effect for Light

39. Show that, for the 21.1-cm line so much used by radio astronomers, a Doppler frequency shift in kHz can be converted to a radial velocity in km/s by multiplying by 0.211, provided that $u \ll c$.

40. Show that, for speeds $u \ll c$, the Doppler shift can be written in the approximate form

$$\frac{\Delta\lambda}{\lambda} = \frac{u}{c},$$

where $\Delta\lambda$ is the change in wavelength.

41. A rocket ship is receding from Earth at a speed of 0.20*c*. A light in the rocket ship appears blue to passengers on the ship. What color would it appear to be to an observer on Earth? (See Fig. 39-6.)

42. A spaceship, moving away from Earth at a speed of 0.892*c*, reports back by transmitting on a frequency (measured in the spaceship frame) of 100 MHz. To what frequency must Earth receivers be tuned to receive these signals?

43. In the spectrum of quasar 3C9, some of the familiar hydrogen lines appear but they are shifted so far toward the red that their wavelengths are observed to be three times as large as that observed in the light from hydrogen atoms at rest in the laboratory. (*a*) Show that the classical Doppler equation, which assumes that light behaves like sound, gives a velocity of recession greater than *c*. (*b*) Assuming that the relative motion of 3C9 and the Earth is entirely one of recession, find the recession speed predicted by the relativistic Doppler equation.

44. The "red shift" of radiation from a distant galaxy consists of the light H_γ, known to have a wavelength of 434 nm when observed in the laboratory, appearing to have a wavelength of 462 nm. (*a*) What is the speed of the galaxy in the line of sight relative to the Earth? (*b*) Is the galaxy approaching or receding?

45. Calculate the Doppler wavelength shifts expected for light of wavelength 553 nm emitted from the edge of the Sun's disk at the equator due to the Sun's rotation. See Appendix C for needed data.

46. With what speed would you have to go through a red light in order to have it appear green? Take 620 nm as the wavelength of red light and 540 nm as the wavelength of green light.

47. In the experiment of Ives and Stilwell, the speed *u* of the hydrogen atoms in a particular run was 8.65×10^5 m/s. Calculate $\Delta f/f$, on the assumptions that (*a*) Eq. 39-18 is correct and (*b*) that Eq. 39-20 is correct; compare your results with those given in Table 39-4 for this speed.

48. *A*, on Earth, signals with a flashlight every six minutes. *B* is on a space station that is stationary with respect to Earth. *C* is on a rocket traveling from *A* to *B* with a constant velocity of 0.60*c* relative to *A*; see Fig. 39-44. (*a*) At what intervals does *B* receive signals from *A*? (*b*) At what intervals does *C* receive signals from *A*? (*c*) If *C* flashes a light every time a flash is received from *A*, at what intervals does *B* receive *C*'s flashes?

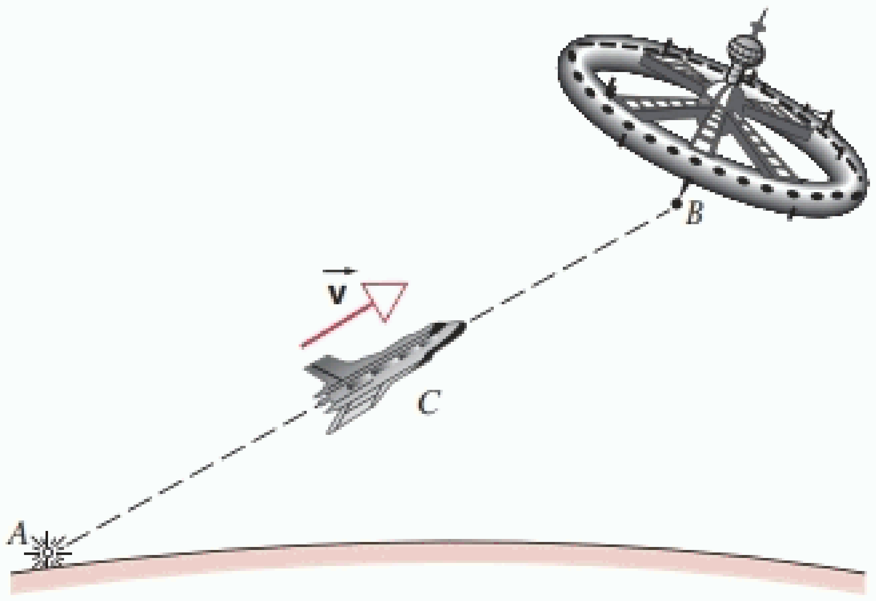


FIGURE 39-44. Exercise 48.

49. An Earth satellite, transmitting on a frequency of 40 MHz, passes directly over a radio receiving station at an altitude of 400 km and at a speed of 2.8×10^4 km/h. Plot the change in frequency, attributable to the Doppler effect, as a function of time, counting $t = 0$ as the instant the satellite is over the station. (Hint: The speed in the Doppler formula is not the actual speed of the satellite but its component in the direction of the station. Neglect the curvature of the Earth and of the satellite orbit.)

50. Give the Doppler wavelength shift $\lambda - \lambda_0$, if any, for the sodium D_2 line (589.00 nm) emitted from a source moving in a circle with constant speed 0.122*c* as measured by an observer fixed at the center of the circle.

51. A source of radio waves, rest frequency 188 MHz, is moving at a speed of 0.717*c* transverse to the line of sight to a detector. At what angle to the line of sight must a second source, rest frequency 162 MHz, move also at 0.717*c* if the frequencies of the two sources as received at the detector are to be equal?

52. A source of light moves at right angles to the line of sight to a detector. The speed of the source is 0.662*c*. At what speed must an identical source move at 75.0° to the line of sight if the Doppler shifts as recorded by the detector for the two sources are equal?

PROBLEMS

1. Consider a star located on a line through the Sun, drawn perpendicular to the plane of the Earth's orbit about the Sun. The star's distance is much greater than the diameter of the Earth's orbit. Show that, due to the finite speed of light, a telescope through which the star is seen must be tilted at an angle $\alpha = 20.5''$ to the perpendicular, in the direction the Earth is moving; see Fig. 39-45. This phenomenon, called aberration, is noticeable and was first explained by James Bradley in 1729.

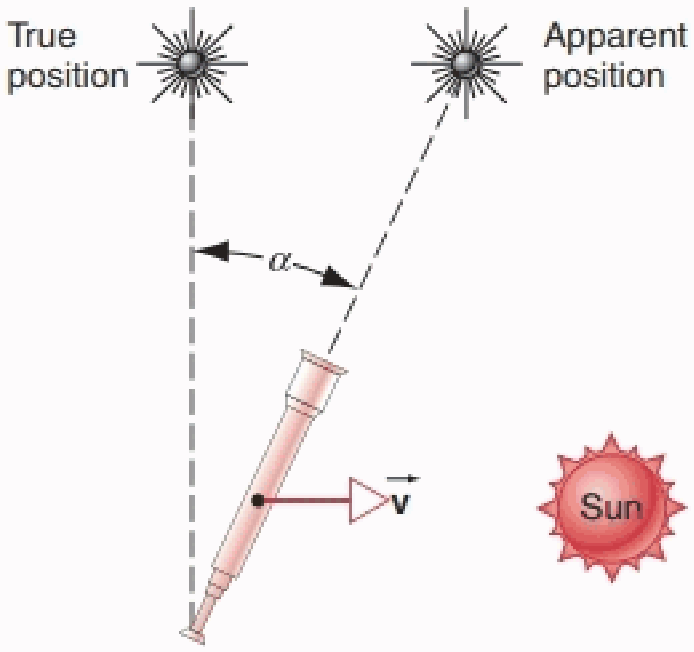


FIGURE 39-45. Problem 1.

2. In 1676, Ole Roemer deduced that the speed of light is finite by observing the time of the eclipse of one of Jupiter's satellites, Io (see Fig. 39-46). Based on the known orbital properties of Io, it was predicted to emerge from Jupiter's shadow at a particular time, corresponding to the Earth at position x in its orbit. When the Earth was actually at position y, Io emerged from Jupiter's shadow about ten minutes late. Roemer concluded that the discrepancy must be due to the additional time necessary for light from Io to travel the additional distance of the radius of the Earth's orbit. What value can be calculated for the speed of light from this observation? (These observations can also be interpreted in terms of the Doppler effect of light. See "The Doppler Interpretation of Roemer's Method," by V. M. Babovicic, D. M. Davidovic, and B. A. Anicin, *American Journal of Physics,* June 1991, p. 515.)

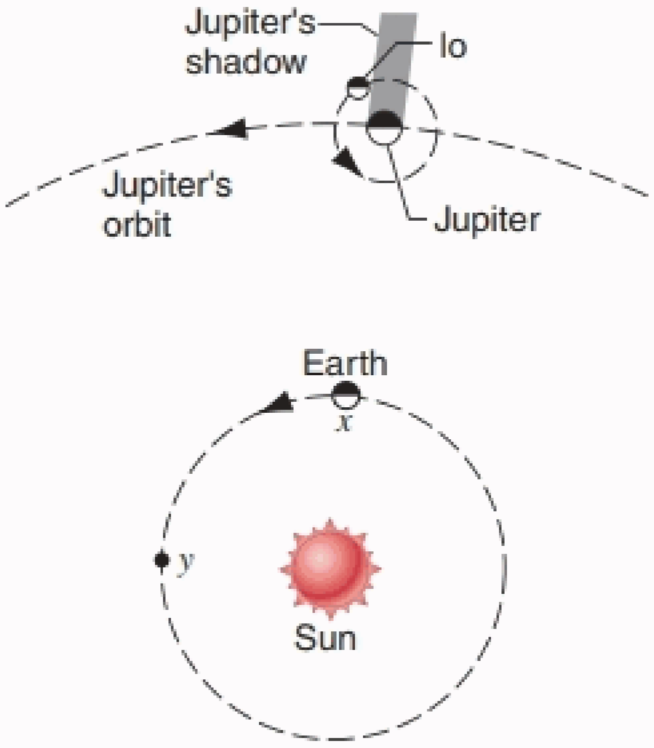


FIGURE 39-46. Problem 2.

3. Ocean waves moving at a speed of 4.0 m/s are approaching a beach at an angle of 30° to the normal, as shown in Fig. 39-47. Suppose the water depth changes abruptly and the wave speed drops to 3.0 m/s. Close to the beach, what is the angle θ between the direction of wave motion and the normal? (Assume the same law of refraction as for light.) Explain why most waves come in normal to a shore even though at large distances they approach at a variety of angles.

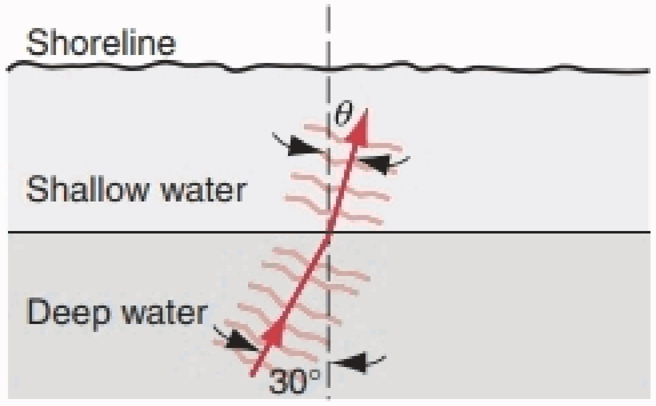


FIGURE 39-47. Problem 3.

4. Two perpendicular mirrors form the sides of a vessel filled with water, as shown in Fig. 39-48. A light ray is incident from above, normal to the water surface. (*a*) Show that the emerging ray is parallel to the incident ray. Assume that there are two reflections at the mirror surfaces. (*b*) Repeat the analysis for the case of oblique incidence, with the ray lying in the plane of the figure.

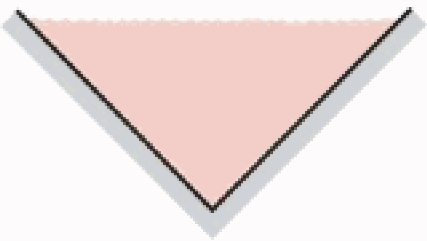


FIGURE 39-48. Problem 4.

5. Prove that a ray of light incident on the surface of a sheet of plate glass of thickness t emerges from the opposite face parallel to its initial direction but displaced sideways, as in Fig. 39-49. (*a*) Show that, for small angles of incidence θ, this displacement is given by

$$x = t\theta\frac{n-1}{n},$$

where n is the index of refraction and θ is measured in radians. (*b*) Calculate the displacement at a 10° angle of incidence through a 1.0-cm-thick sheet of crown glass.

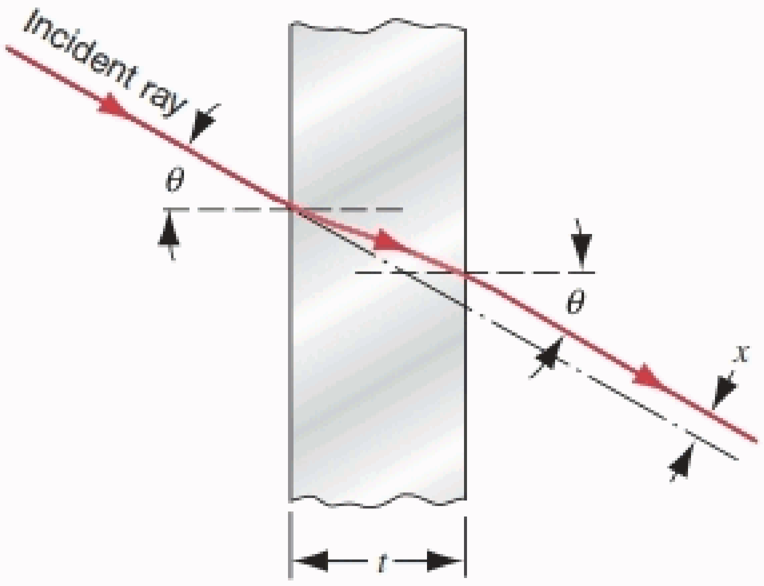


FIGURE 39-49. Problem 5.

6. The index of refraction of the Earth's atmosphere decreases monotonically with height from its surface value (about 1.00029) to the value in space (about 1.00000) at the top of the atmosphere. This continuous (or graded) variation can be approximated by considering the atmosphere to be composed of three (or more) plane parallel layers, in each of which the index of refraction is constant. Thus, in Fig. 39-50,

$$n_3 > n_2 > n_1 > 1.00000.$$

Consider a ray of light from a star S that strikes the top of the atmosphere at an angle θ with the vertical. (*a*) Show that the apparent direction θ_3 of the star with the vertical as seen by an observer at the Earth's surface is obtained from

$$\sin \theta_3 = \frac{1}{n_3} \sin \theta.$$

(Hint: Apply the law of refraction to successive pairs of layers of the atmosphere; ignore the curvature of the Earth.) (*b*) Calculate the shift in position of a star observed to be 50° from the vertical. (The very small effects due to atmospheric refraction can be most important; for example, they must be taken into account in using navigation satellites to obtain accurate fixes of position on the Earth.)

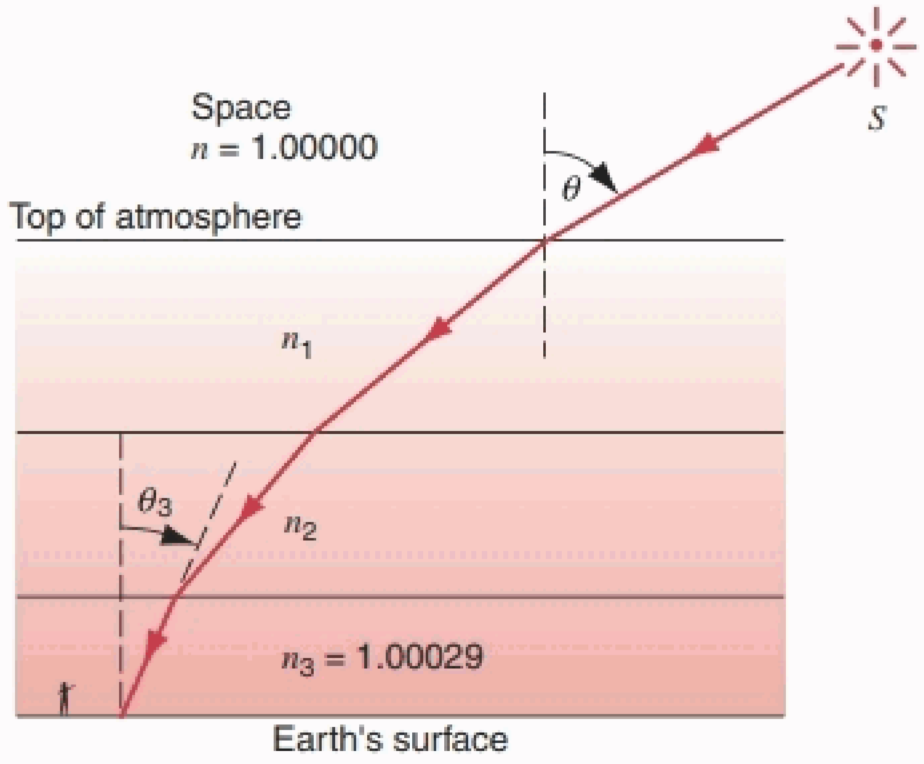


FIGURE 39-50. Problem 6.

7. You stand at one end of a long airport runway. A vertical temperature gradient in the air has resulted in the index of refraction of the air above the runway to vary with height y according to $n = n_0(1 + ay)$, where n_0 is the index of refraction at the runway surface and $a = 1.5 \times 10^{-6}\ \text{m}^{-1}$. Your eyes are at a height $h = 1.7$ m above the runway. Beyond what horizontal distance d can you not see the runway? See Fig. 39-51 and Problem 6.

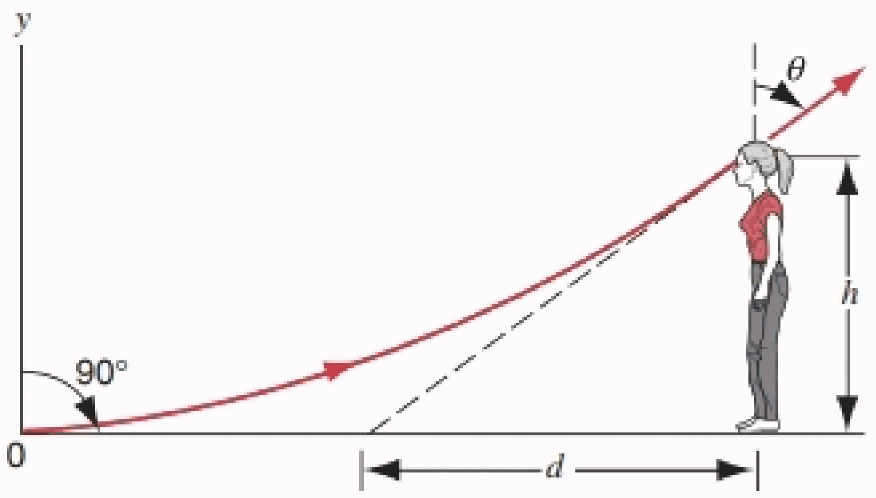


FIGURE 39-51. Problem 7.

8. Muons (mass = 106 MeV/c^2) and neutral pions (mass = 135 MeV/c^2), each with momentum 145 MeV/c, pass through a transparent material. Find the range of index of refraction of the material so that only the muons emit Cerenkov radiation. (See Exercise 17.)

9. A drop of liquid may be placed on a semicircular slab of glass as in Fig. 39-52. (*a*) Show how to determine the index of refraction of the liquid by observing total internal reflection. The index of refraction of the glass is unknown and must also be determined. Is the range of indices of refraction that can be measured in this way restricted in any sense? (*b*) In reality, how practical is this method?

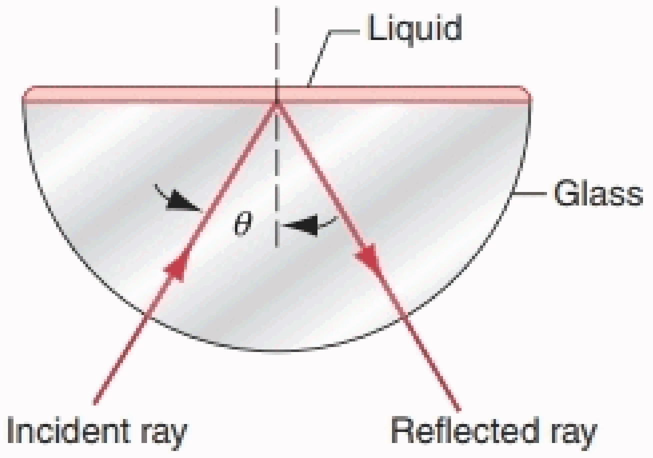


FIGURE 39-52. Problem 9.

10. A light ray of given wavelength, initially in air, strikes a 90° prism at P (see Fig. 39-53) and is refracted there and at Q to such an extent that it just grazes the right-hand prism surface at Q. (*a*) Determine the index of retraction of the prism for this wavelength in terms of the angle of incidence θ_1 that gives rise to this situation. (*b*) Give a numerical upper bound for the index of refraction of the prism. Show, by ray diagrams, what happens if the angle of incidence at P is (*c*) slightly greater or (*d*) slightly less than θ_1.

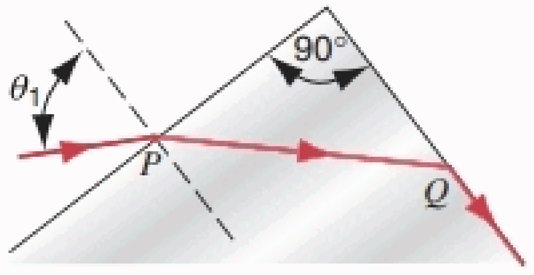


FIGURE 39-53. Problem 10.

11. A point source of light is placed a distance h below the surface of a large, deep lake. (*a*) Show that the fraction f of the light energy that escapes directly from the water surface is independent of h and is given by

$$f = \frac{1}{2}(1 - \sqrt{1 - 1/n^2}),$$

where n is the index of refraction of water. [Note: Absorption within the water and reflection at the surface (except where it is total) have been neglected.] (*b*) Evaluate this fraction numerically.

12. A particular optical fiber consists of a nongraded glass core (index of refraction n_1) surrounded by a cladding (index of refraction $n_2 < n_1$). Suppose a beam of light enters the fiber from air at an angle θ with the fiber axis as shown in Fig. 39-54. (*a*) Show that the greatest possible value of θ for which a ray can be propagated down the fiber is given by

$$\theta = \sin^{-1}\sqrt{n_1^2 - n_2^2}.$$

(*b*) Assume that the glass and coating indices of refraction are 1.58 and 1.53, respectively, and calculate the value of this angle.

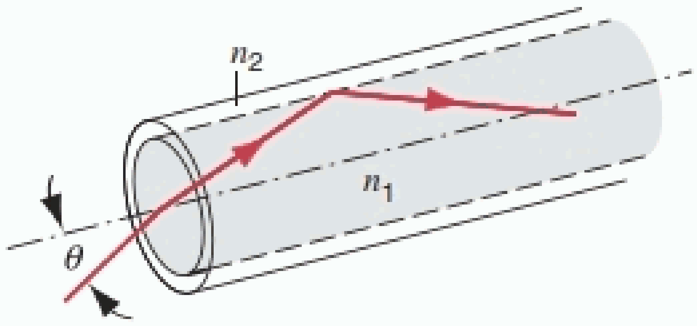


FIGURE 39-54. Problem 12.

13. In an optical fiber (see Problem 12), different rays travel different paths along the fiber, leading to different travel times. This causes a light pulse to spread out as it travels along the fiber, resulting in information loss. The delay time should be minimized in designing a fiber. Consider a ray that travels a distance L along a fiber axis and another that is reflected, at the critical angle, as it travels to the same destination as the first. (*a*) Show that the difference Δt in the times of arrival is given by

$$\Delta t = \frac{L}{c}\frac{n_1}{n_2}(n_1 - n_2),$$

where n_1 is the index of refraction of the core and n_2 is the index of refraction of the cladding. (*b*) Evaluate Δt for the fiber of Problem 12, with $L = 350$ km.

14. Hydrogen molecules at 700 K emit light of frequency 457 THz. (*a*) Determine the change in frequency of the light observed due to the motion of a molecule moving toward an observer with the root-mean-square speed. (*b*) Find the frequency shift if such light originated from hydrogen atoms instead of molecules.

15. Microwaves, which travel with the speed of light, are reflected from a distant airplane approaching the wave source. It is found that, when the reflected waves are beat against the waves radiating from the source, the beat frequency is 990 Hz. If the microwaves are 12.0 cm in wavelength, what is the approach speed of the airplane?

COMPUTER PROBLEM

1. A rainbow occurs because white light refracts and reflects off of a spherical droplet as shown in Fig. 39-55. (*a*) Generate a graph that shows the angle of the exiting light as a function of the impact parameter b. (*b*) Find the angular limits ϕ_1 and ϕ_2 between which the exiting rays contain 80% of the incident light.

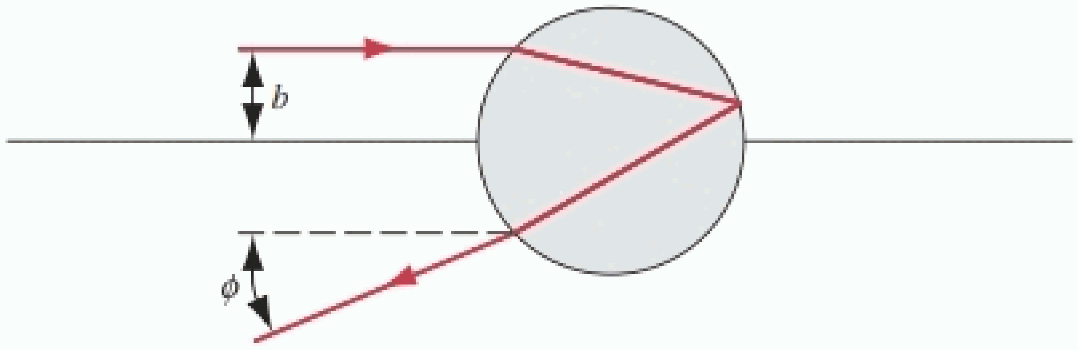


FIGURE 39-55. Computer problem 1.

CHAPTER 40

MIRRORS AND LENSES

Optical systems containing mirrors and lenses are of great practical importance. We can use such systems to correct faulty vision, project an image on a screen where it can be viewed by many people at the same time (such as in a movie theater or in your classroom), and make small objects appear large (as in a microscope) or distant objects appear close (as in a telescope).

In this chapter we study the formation of images by mirrors and lenses. We will develop both algebraic and graphical methods for analyzing image formation, and we will extend these methods to systems with two or more components, such as microscopes or telescopes.

40-1 IMAGE FORMATION BY MIRRORS AND LENSES

When you use a mirror or a lens to look at your friend, you will see something that is not really there. Your friend is very real, of course, but the mirror or lens distorts what you see, so that it may look very different from your actual friend in location (closer to you or farther away), size (your friend may appear larger or smaller), and perhaps even orientation (your friend might appear upside-down).

What you are seeing is an *image* of your friend. To understand the formation of an image, we must first understand how the path of a light beam is changed when it encounters the mirror or lens, for which we use the laws of reflection and refraction that we developed in Chapter 39. However, we must also understand how our brain processes the light that our eyes observe coming from the mirror or lens. Our brain has a very strong desire to believe that light travels only in straight lines. That is why, when you look into a plane mirror, you think that your image is located somewhere *behind* the mirror.

When you see an object, your brain is responding to light that enters your eyes from the object. Your brain takes that information about the object, adds a few clues from the environment, pulls up a few facts from your memory, and produces an image of the object and its surroundings. If any of these ingredients is inconsistent with the others, your brain does the best job it can to make sense of it all. Sometimes, as in the case of "optical illusions," the brain can be fooled completely.

The process is similar when you view light coming from a mirror or lens. Your brain tries to process the information into the most consistent interpretation it can make. In the case of your image in a plane mirror, for example, your brain seems to want to place you somewhere behind the wall to which your mirror is attached!

The type of image that is formed by a plane mirror, and also in some cases by curved mirrors and lenses, is called a *virtual* image. It is characterized by several properties: (1) No light is actually passing through the apparent location of the image. In fact, as in the case of the plane mirror, the image might appear in a location where the light cannot possibly travel. (2) The image cannot be focused on a screen. To see the image, you must look *into* the mirror or the lens. (3) A virtual image produced by a single mirror or lens is always upright (not inverted), although (as we discuss at the end of this chapter) optical systems with two or more lenses or mirrors can produce virtual images that may be either upright or inverted.

A *real* image, on the other hand, has a different set of properties. (1) After encountering the mirror or lens, the light passes directly through the location of the image. (2) The image, therefore, *can* be focused on a screen. To see the image, it is not necessary to look into the mirror or lens. (Think of a movie projector or the overhead projector in your classroom; you can view the image on the screen without actually looking into the lens.) (3) A real image produced by a single mirror or lens is always inverted (upside-down), but again systems with several components can produce real images that may be either upright or inverted.

As we develop the rules for analyzing image formation by mirrors and lenses, we will expand our discussion about the conditions under which real or virtual images may form. In the meantime, you can try the following experiment: Hold a shiny metal spoon about a foot from your face so that the hollow bowl of the spoon is facing you. You should see an inverted view of your face; that is a real image, which is actually located in space about a centimeter in front of the spoon. Now turn the spoon over and look at your image in the back side. You should see an upright view of your face; that is the virtual image, which is located (as we shall discuss) behind the spoon.

Geometrical vs. Wave Optics

Figure 40-1 shows a top view of a water tank in which waves are produced at the left side and travel from left to right. The bright and dark bands in the photograph represent crests and valleys of the waves; thus adjacent bright bands (or dark bands) are one wavelength apart. The waves encounter a barrier into which a small hole has been cut. You can see how the waves flare out after they pass through the hole and appear in the region behind the solid barrier. This phenomenon is known as *diffraction.* Diffraction, which we discuss in detail in Chapter 42, occurs whenever a wave encounters an opening whose dimensions are roughly of the same size as (or smaller than) the wavelength of the wave. In the case of Fig. 40-1, the size of the opening is about the same as the wavelength. The flaring effect decreases as the size of the opening is made larger than the wavelength. If the size of the opening is many times the wavelength, the effects of diffraction are negligible, and after passing through the opening the wave will appear only in the geometrical "shadow" of the opening. Similar effects occur if the wave encounters an obstacle instead of an opening in a barrier.

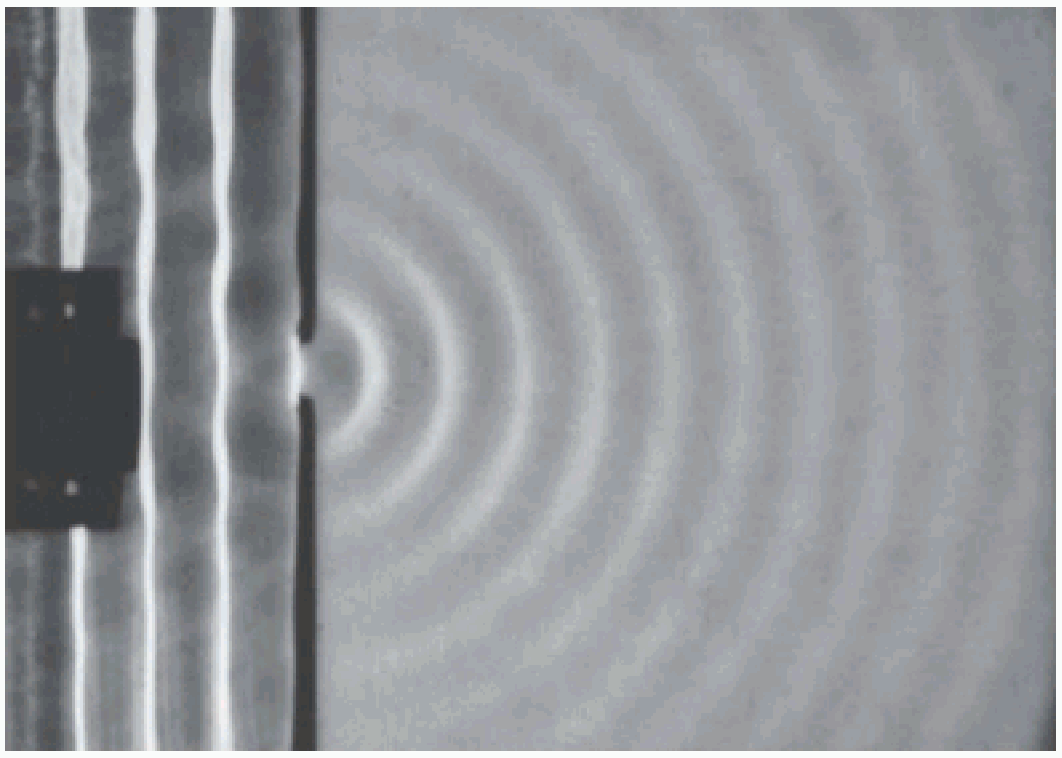


FIGURE 40-1. Diffraction of water waves at an opening in a barrier. Note that the opening is about the same size as the wavelength.

The wavelength of visible light is in the range of 400 to 700 nm (4×10^{-7} to 7×10^{-7} m). Even a tiny mirror or lens of diameter only a few mm is 4 orders of magnitude larger than the wavelength of light. Diffraction effects are thus almost always negligible in the formation of images by mirrors and lenses. This is the realm of *geometrical optics,* in which light beams encounter objects that are much larger than the wavelength of the light. The conditions of geometrical optics will apply in all cases that we discuss in this chapter.

In geometrical optics we can analyze the formation of images by assuming the light to travel in straight-line paths or *rays.* For this reason, geometrical optics is also known as *ray optics.* A ray is a convenient way of describing the travel of light; rays are drawn perpendicular to the wavefronts in a direction that represents the travel of the wave. We have in fact already used rays in Section 39-4 to represent light as it undergoes reflection and refraction.

On the other hand, if diffraction effects occur, we are in the realm of *physical optics,* also known as *wave optics* because we must take the wave nature of light specifically into account in the analysis. Interference (which we discuss in Chapter 41) and diffraction (Chapter 42) are examples of effects for which we must use the methods of physical optics.

40-2 PLANE MIRRORS

Perhaps our most familiar optical experience is looking into a mirror. Figure 40-2 shows a point source of light O, which we call the *object,* placed a distance o in front of a plane mirror. The light that falls on the mirror is represented by rays emanating from O.* At the point at which each ray strikes the mirror we construct a reflected ray, using the law of reflection (Eq. 39-3). If we extend the reflected rays backward, they intersect at a point I, which we call the *image* of the object O. The image is the same distance behind the mirror that the object O is in front of it, which we prove below.

---

*In our previous discussion of reflection in Chapter 39, we assumed an incident *plane* wave; the incident rays are parallel to each other in that case. Here we have a *point* source, and the rays striking the mirror are *diverging* from that point source. We can regard the point source as a source of spherical waves, and the rays radiating from the source are perpendicular to the spherical wavefronts.

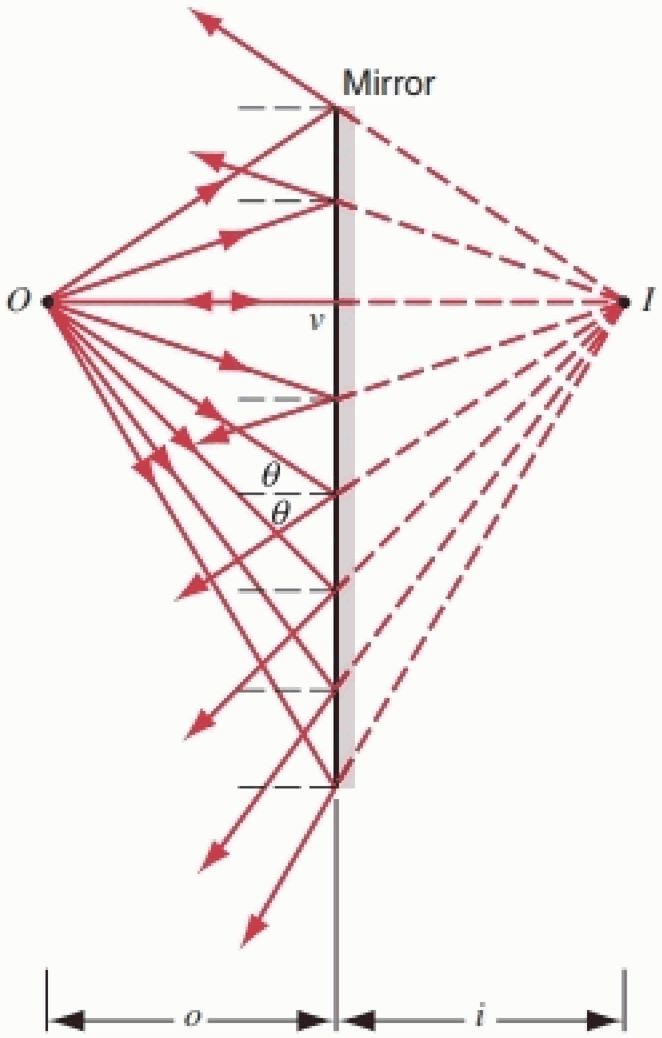


FIGURE 40-2. A point object O forms a virtual image I in a plane mirror. The rays *appear* to diverge from I but no light is actually present at that point.

Images of diverging light in plane mirrors are always virtual. We know from daily experience how "real" such a virtual image *appears to be* and how definite is its location in the space behind the mirror, even though this space may, in fact, be occupied by a brick wall.

Figure 40-3 shows two rays from Fig. 40-2. One strikes the mirror at v, along a perpendicular line. The other strikes it at an arbitrary point a, making an angle of incidence θ with the normal at that point. Elementary geometry shows that the angles aOv and aIv are also equal to θ. Thus the right triangles aOv and aIv are congruent and so

$$i = -o, \tag{40-1}$$

in which we introduce the negative sign to signify that I and O are on opposite sides of the mirror. Equation 40-1 does not involve θ, which means that *all* rays from O striking the mirror pass through I when extended backward, as we have seen in Fig. 40-2. Other than assuming that the mirror is truly plane and that the conditions for geometrical optics hold, we have made no approximations in deriving Eq. 40-1. A point object produces a point image in a plane mirror, with $i = -o$, no matter how large is the angle θ in Fig. 40-3.

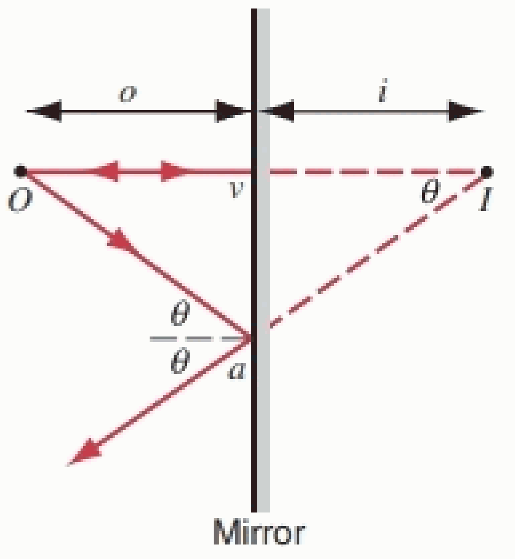


FIGURE 40-3. Two rays from Fig. 40-2. Ray Oa makes an arbitrary angle θ with the normal to the surface of the mirror.

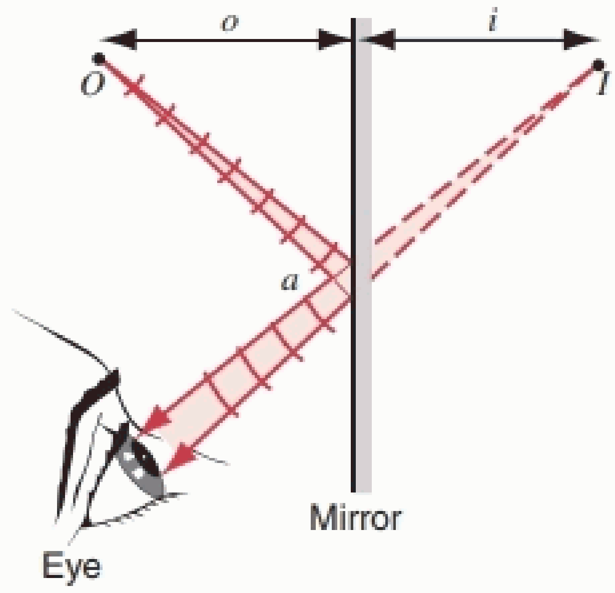


FIGURE 40-4. A pencil of rays from O enters the eye after reflection from the mirror. Only a small portion of the mirror near a is effective. The small arcs represent portions of the spherical wavefronts. The light appears to come from I.

Because of the finite diameter of the pupil of the eye, only rays that lie fairly close together can enter the eye after reflection at a mirror. For the eye position shown in Fig. 40-4, only a small patch of the mirror near point a is effective in forming the image; the rest of the mirror may be covered up or removed. If we move our eye to another location, a different patch of the mirror will be effective; the location of the virtual image I will remain unchanged, however, as long as the object remains fixed.

The Image of an Extended Object

Often we are interested not in the image of a single point source of light, but of an extended object, such as a face or a full body. As in Fig. 40-5, we represent the extended object by an arrow. All points on the object are the same distance o from the mirror. Using the methods of this section, we can show that object point O forms an image I that is a distance i behind the mirror. The object and image distances are related by Eq. 40-1 ($i = -o$). Object point O' forms image I'

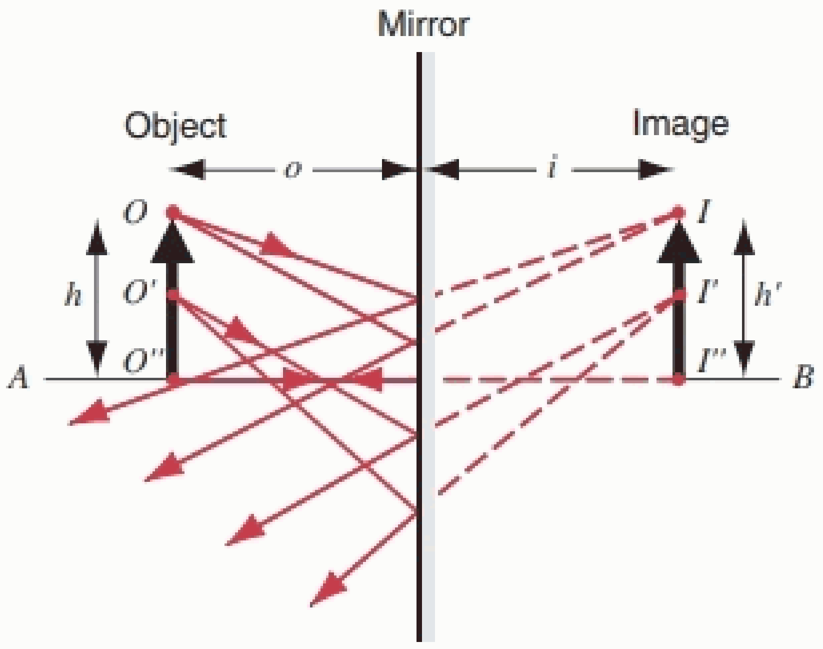


FIGURE 40-5. The image of an extended object in a plane mirror.

that is also a distance i behind the mirror. By considering the entire object to be composed of point sources, we could eventually trace out the entire profile of the image.

Because all points on the object are a distance o from the mirror and all points on the image are a distance i behind the mirror, *it is necessary to find the location of only one image point to locate the entire image.* Usually we find the image of the tip of the object, which then locates the tip of the image. It is then easy to draw the entire image, because the image I'' of the tail of the arrow must lie along the line AB that extends from the tail of the object O'' and is perpendicular to the mirror.

For an extended object, it is convenient to define the *lateral magnification m* (which we usually refer to simply as the magnification) to be the ratio of the height h' of the image to the height h of the object:

$$m = \frac{h'}{h}. \tag{40-2}$$

We adopt the convention that heights measured above the line AB (which we will later call the *axis* of the mirror or lens) are positive and those below AB are negative. If the image is upright with respect to the object (that is, both object and image extend in the same direction from the axis, either both up or both down), the magnification is positive; if the image is inverted with respect to the object (one up and the other down), the magnification is negative. Later we will show that a general equation for the magnification, which also applies to a curved mirror or lens, is

$$m = -\frac{i}{o}. \tag{40-3}$$

For the plane mirror, both Eqs. 40-2 and 40-3 show that $m = +1$; that is, the image is upright with respect to the object (indicated by the + sign), and it is the same size as the object (because the magnitude of the magnification is 1). If $|m| > 1$, the image is enlarged relative to the object; if $|m| < 1$, the image is reduced.

Image Reversal

As Fig. 40-6*a* shows, the image of a left hand appears to be a right hand. We interpret this appearance as a reversal of left and right. That is, if you raise your left hand, then your mirror image raises its right hand. It is often then asked: Why does a mirror reverse left and right but not also reverse up and down?

Figure 40-6*b* illustrates the way a mirror reverses the image of a three-dimensional object, represented simply as a set of three mutually perpendicular arrows. Note that the arrows parallel to the plane of the mirror (arrows x and y) are identical with their mirror images. Only the z arrow has its direction changed by the reflection. It is therefore more accurate to say that a mirror reverses front to back rather than left to right. The transformation of a left hand to a right hand is accomplished, in a sense, by exchanging the front and back of the hand.

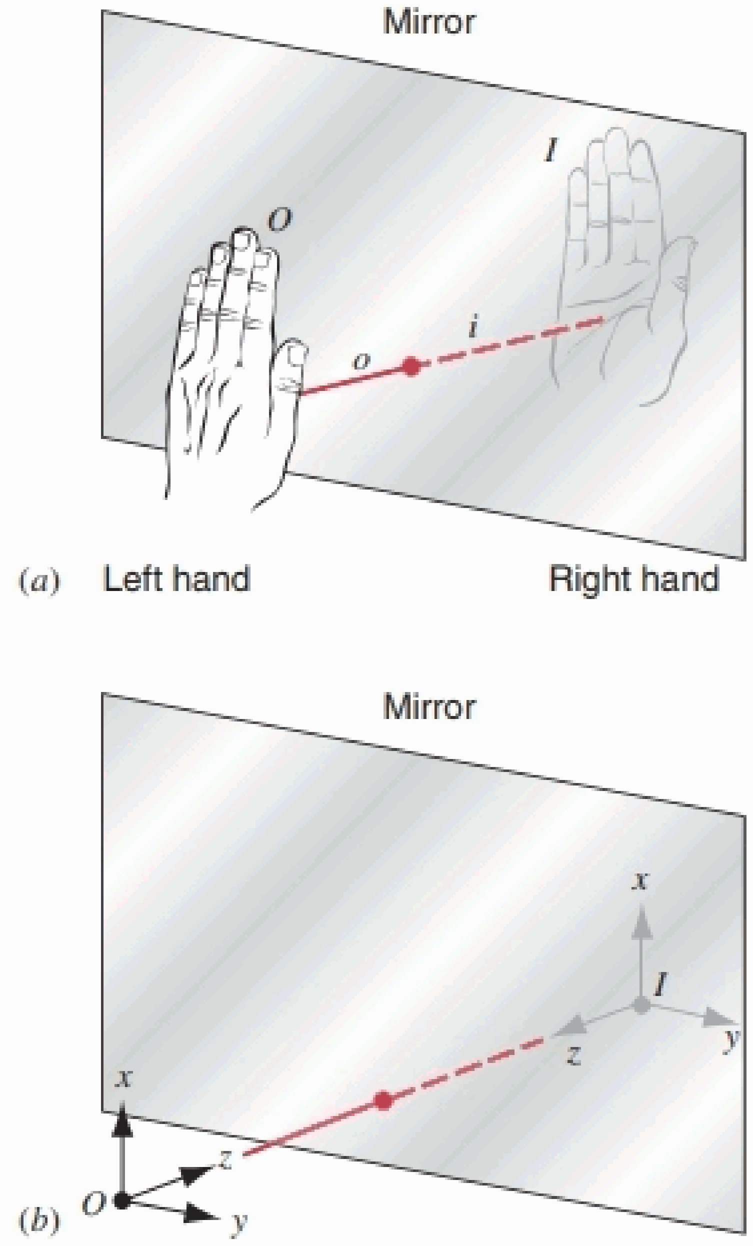


FIGURE 40-6. (*a*) The object O is a left hand; its image I is a right hand. (*b*) Study of a reflected three-arrow object shows that a mirror interchanges front and back, rather than left and right.

Note also that the object can be considered to represent a conventional right-handed coordinate system (x "crossed into" y points in the z direction), while the image is a left-handed coordinate system (x "crossed into" y points in the *negative* z direction). Such reversals apply to physical objects as well; for example, the image of a screw with right-handed threads is a screw with left-handed threads.

If we knew for a fact that all humans were right-handed, then we could surely tell the difference between a physical situation and its mirror image; the "real" person would be using the right hand, whereas the image would use the left. However, if humans were ambidextrous, we could not use this feature to distinguish between the real world and the looking-glass world. The same distinction has been applied to the laws of physics: if the laws of physics have perfect left–right symmetry, then the mirror image of an experiment would also be a physically possible experiment. If, however, the laws lack that symmetry, then the outcome of some mirror-image experiments would not be physically possible. In 1956, it was discovered that the so-called weak interaction, which causes certain radioactive decays, lacks this symmetry, which is called *parity.* This experiment provided the first fundamental basis for a distinction between our world and its mirror image.*

---

*For some amusing discussions about symmetry and the distinctions between objects and their mirror images, see *The Ambidextrous Universe,* by Martin Gardner (Scribner's, 1979), and *Reality's Mirror,* by Bryan Bunch (Wiley, 1989).

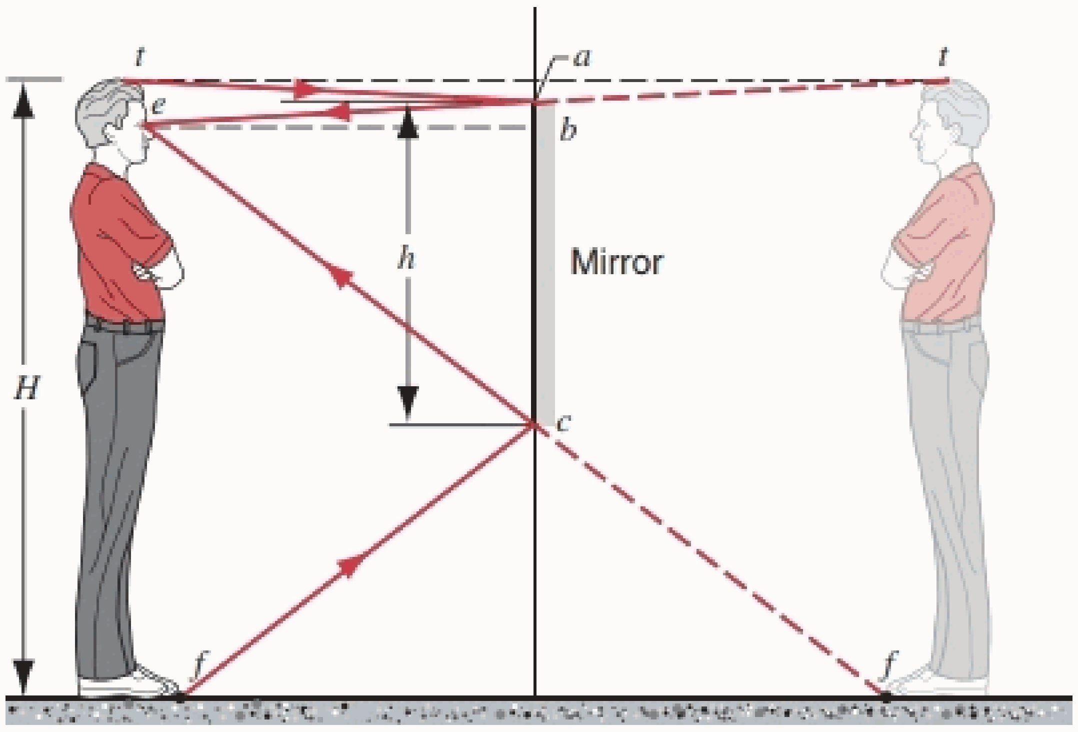


FIGURE 40-7. Sample Problem 40-1.

SAMPLE PROBLEM 40-1. Find the minimum length h of a mirror that is needed for a person of height H to see his entire reflection.

Solution Figure 40-7 shows the person's foot f, eyes e, and top of head t. For him to see his entire height, a ray of light (tae) must leave the top of his head, reflect from the mirror at a, and enter his eyes, while another ray (fce) must leave his feet, reflect from the mirror at c, and enter his eyes. The person will see a full-height reflection (including the virtual images of points t and f) if the length of the mirror is at least ac.

From the geometry of Fig. 40-7, we see that

$$ab = \tfrac{1}{2}te \quad \text{and} \quad bc = \tfrac{1}{2}ef,$$

where point b is at the same height as the eyes. Thus

$$ac = ab + bc = \tfrac{1}{2}te + \tfrac{1}{2}ef = \tfrac{1}{2}tf.$$

With $h = ac$ and $H = tf$, we obtain

$$h = \tfrac{1}{2}H.$$

The person can see his entire image if the mirror is at least half his height. Portions of the mirror below c show reflections of the floor in front of his feet, while those above t show the person what is above his head. Note that the distance of the person from the mirror makes no difference in this calculation, which remains valid for any object distance from a plane mirror.

40-3 SPHERICAL MIRRORS

Suppose that, instead of making the mirror flat, we give it a slight curvature. In particular, we consider mirrors that have a spherical shape. Figures 40-8*b* and 40-8*c* show the effect in two different cases. In the first case (Fig. 40-8*b*), the mirror is *concave* (meaning "hollow," like a cave) with respect to the location of the object. Note that, in comparison with the plane mirror, the image is (1) magnified (that is, larger than the object) and (2) located at a greater distance behind the mirror (that is, i has a larger negative value). Such mirrors are commonly used for shaving or applying make-up, when magnification is desirable, even though the field of view may be reduced. Figure 40-8*b* applies only when the distance of the object from the mirror is small (less than $r/2$, as we shall see).

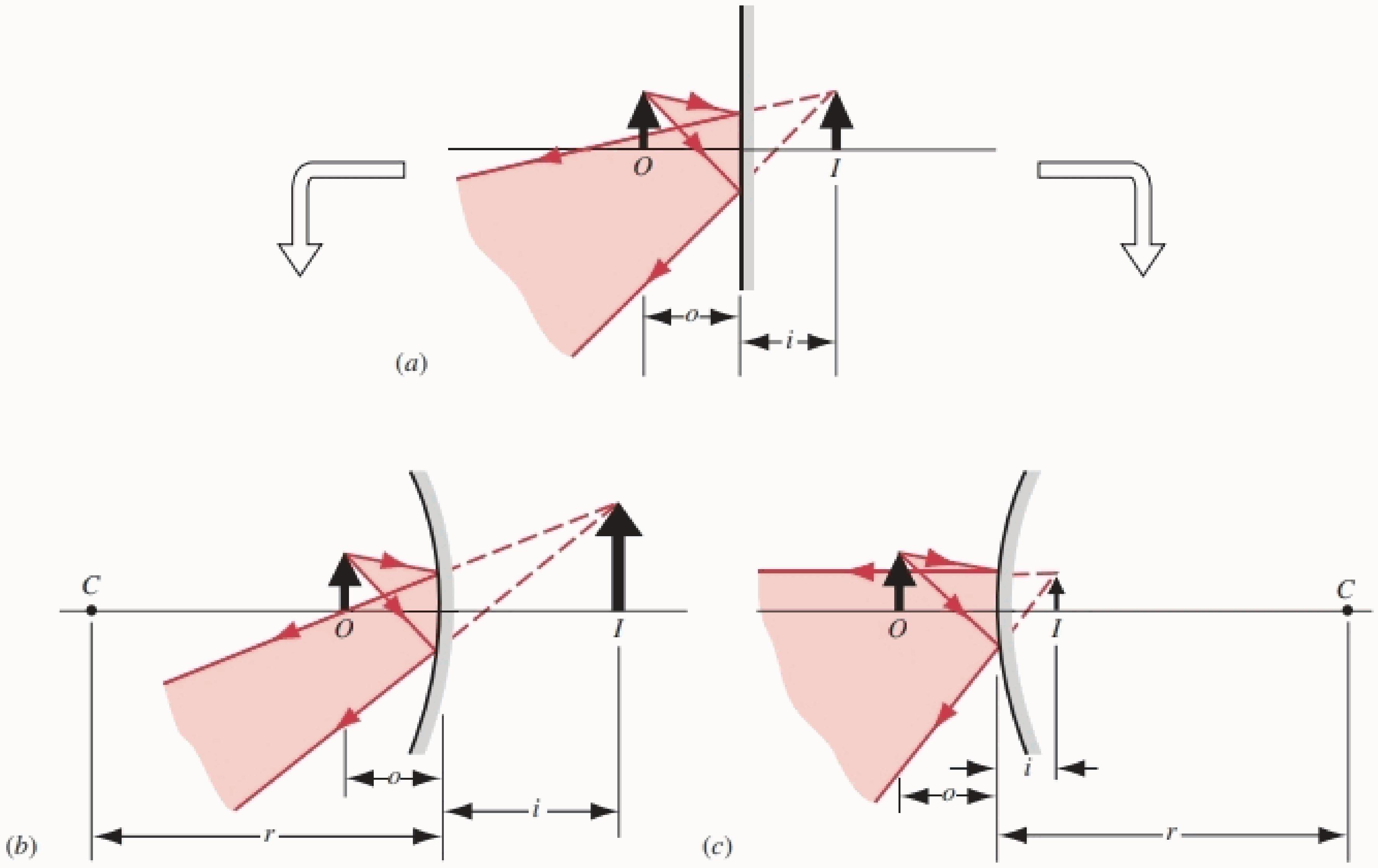


FIGURE 40-8. (*a*) An object O forms a virtual image I in a plane mirror. (*b*) If the mirror is bent so it becomes *concave,* the image moves away from the mirror and becomes larger. (*c*) If the plane mirror is bent so it becomes *convex,* the image moves closer to the mirror and becomes smaller. Point C is called the *center of curvature* of the mirror; it is the center of the spherical surface of which the mirror is a part.

In the second case (Fig. 40-8*c*), the mirror is *convex* with respect to the location of the object. Note that the image is (1) reduced in size and (2) closer to the mirror, compared with the plane mirror. Examples of such mirrors are right-hand side-view mirrors in automobiles and surveillance mirrors used in retail stores. The field of view is wider than that of a plane mirror.

Suppose the spherical mirrors in Fig. 40-8 were flexible. If we were to bend either one to make it more planar, the image would approach the location and size of the image in a plane mirror. We can therefore consider a plane mirror to be a special case of a spherical mirror, in which the radius of the sphere becomes infinitely large. Our equations describing the spherical mirror should reduce to the plane mirror equation ($i = -o$) as the radius tends toward infinity.

The Mirror Equation

At the end of this section we derive the equation that relates the object distance o and the image distance i for a spherical mirror. We consider the special case in which the rays of light from the object make small angles with the axis of the mirror. Such rays are called *paraxial* rays. Put another way, the dimensions of the mirror are small compared with its radius of curvature. Our description would not apply to a fully illuminated mirror in the shape of an entire hemisphere.

The mirror equation relates the three distances in Fig. 40-8: o, i, and the radius of curvature r of the mirror. This relationship is given by the spherical mirror equation,

$$\frac{1}{o} + \frac{1}{i} = \frac{2}{r}. \qquad (40\text{-}4)$$

It is convenient to define the *focal length* f of the mirror to be half the radius of curvature, or

$$f = r/2. \qquad (40\text{-}5)$$

In terms of the focal length, the mirror equation can be written

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f}. \qquad (40\text{-}6)$$

Figure 40-9 shows parallel light rays incident on the mirror. Parallel light can be obtained from an object at a great distance from the mirror, such that the wavefronts from the object are essentially planes. In practice, parallel light can be obtained using another mirror or a lens. Parallel light is brought to a focus at a point F, called the *focal point*. This point is distance f from the mirror. Equation 40-6 shows that if $o = \infty$, corresponding to the object at a very great distance from the mirror, then $i = f$.

If $r = \infty$, corresponding to a plane mirror, then Eq. 40-4 reduces to $i = -o$ in agreement with Eq. 40-1. Thus a plane mirror is simply a special case of a spherical mirror for an infinite radius of curvature. Equation 40-6 can be used to find the *location* of the image; its size, compared with the size of the object, can be determined from the magnification, $m = -i/o$ (Eq. 40-3).

Sign Conventions

Figure 40-9 suggests the sign conventions that must be considered in using Eqs. 40-4 and 40-6. The side of the mirror from which the light is incident is called the *R*-side, because it is on this side that a *real image* will be formed. Real images are those that are formed by converging light; equivalently, we can say that real images are those that can be viewed on a screen placed at the position of the image. On the *R*-side of the mirror, i, o, r, and f are taken to be positive.

The region behind the mirror is called the *V*-side, because on this side *virtual images* can be formed. Virtual images are those formed by diverging light that cannot be shown on a screen. On the *V*-side, o, i, r, and f are taken to be negative.

According to these sign conventions, in Fig. 40-8*b* the object distance o is positive (because the object is on the *R*-side of the mirror) and the image distance i is negative (because the image is on the *V*-side). The center of curvature C is on the *R*-side, so the radius of curvature r is positive. In Fig. 40-8*c*, o is positive and i is negative, as in Fig. 40-8*b*, but r is negative, because C is on the *V*-side.

Figure 40-10 shows the image distances for three different object distances as an object is moved toward a concave mirror. In Fig. 40-10*a*, the object and image distances are positive, because the object and the image both appear on

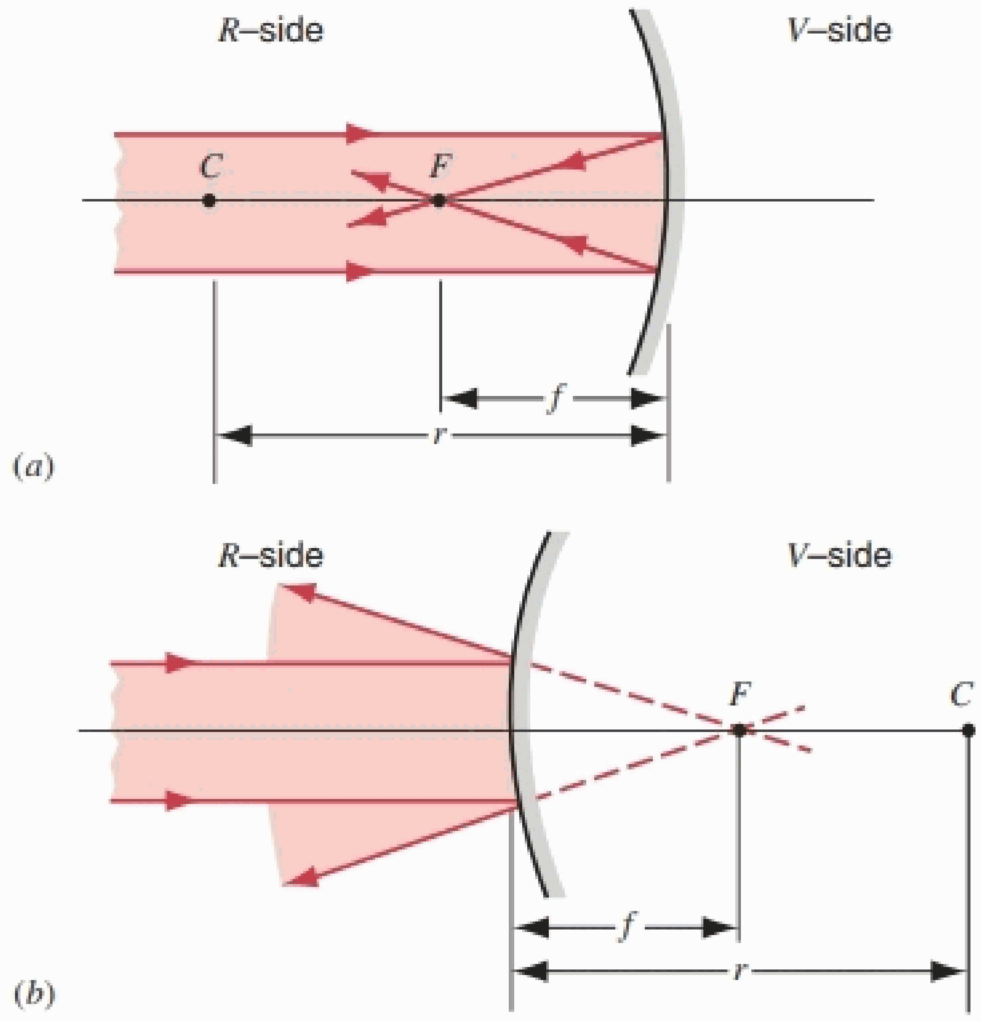


FIGURE 40-9. (*a*) In a concave mirror, incident parallel light is brought to a real focus at F on the *R*-side of the mirror. (*b*) In a convex mirror, incident parallel light appears to diverge from a virtual focus at F on the *V*-side of the mirror.

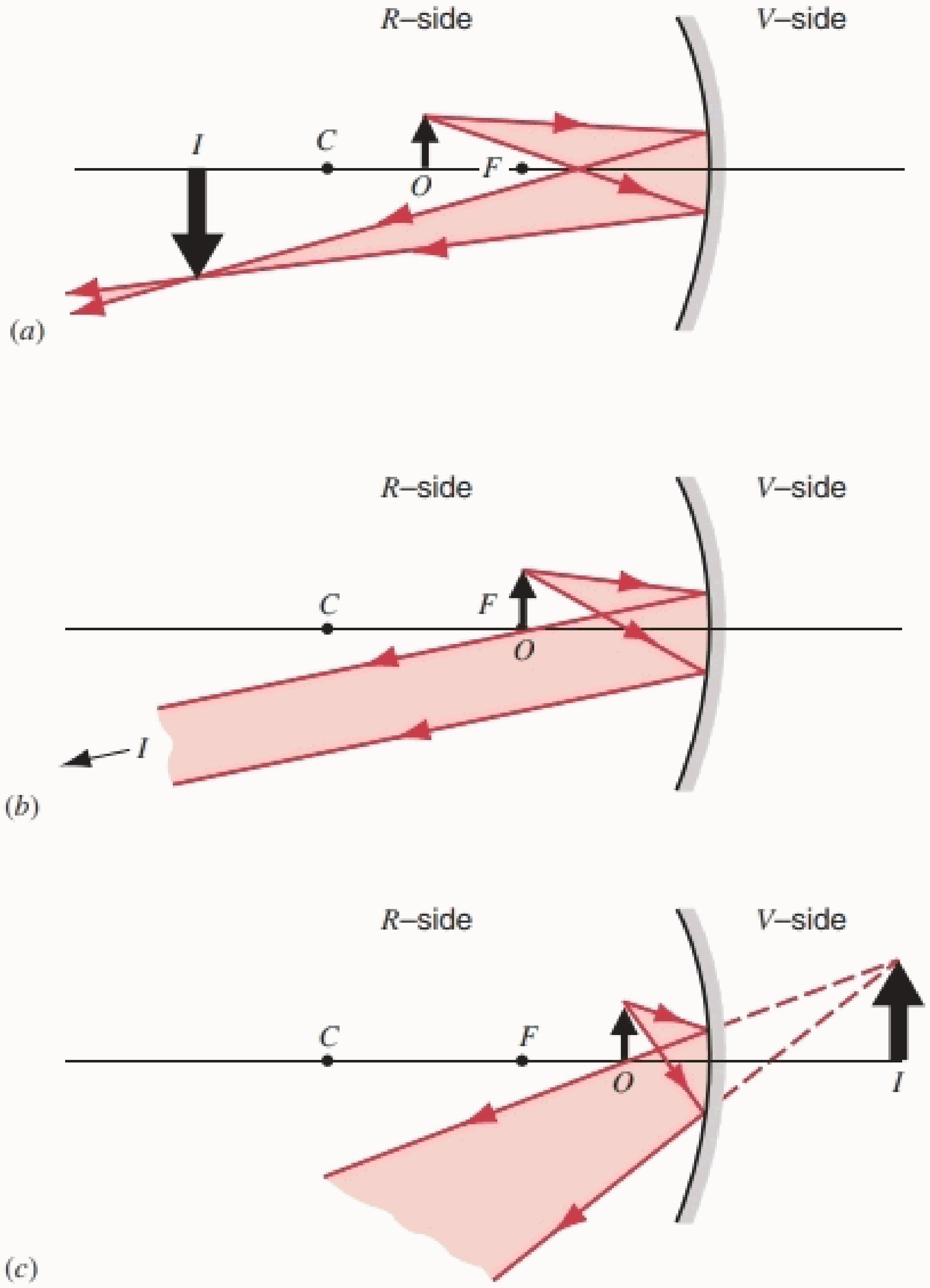


FIGURE 40-10. An object is moved successively closer to a concave mirror, from (*a*) just beyond the focal point to (*b*) the focal point and then (*c*) within the focal point. In the process, the image moves from (*a*) its position on the *R*-side to (*b*) infinity and then (*c*) reappears on the *V*-side.

the *R*-side of the mirror. In Fig. 40-10*b*, the object is at the focal point. With $o = f$, Eq. 40-6 gives $i = \infty$. This is consistent with parallel light emerging from the mirror. In Fig. 40-10*c*, the object distance remains positive but is now smaller than f. In this case, Eq. 40-6 gives a negative value for i; that is, a virtual image forms on the *V*-side, as shown.

In Fig. 40-10*a*, the lateral magnification m determined according to Eq. 40-3 is negative, because o and i are both positive. The image is therefore inverted. (It is also enlarged, because i happens to be greater than o in the case illustrated.) In Fig. 40-10*c*, o and i have opposite signs, so m is positive and the image is correspondingly upright, as illustrated.

Figure 40-11 suggests one possible arrangement in which the object is considered to be on the *V*-side of the mirror, so that o is negative. Converging light (produced by another optical device, such as a lens or a mirror, which is not shown) is incident on the mirror. If the mirror were not present, the light would converge to an image at the location O shown. This location defines the position of a *virtual object,* and the distance between this location and the mirror is the (negative) object distance. The image distance is positive.

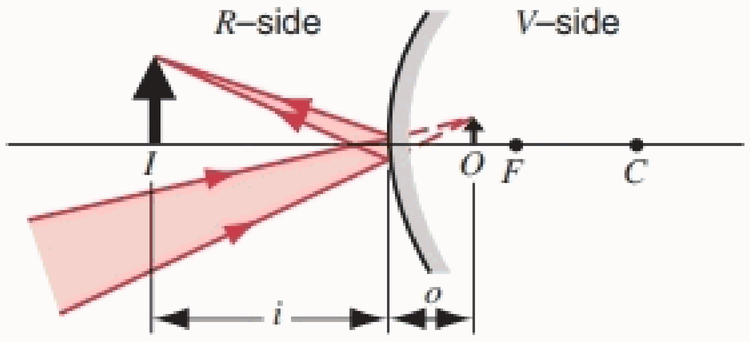


FIGURE 40-11. Converging light (from mirrors or lenses not illustrated) is incident on a convex mirror. The *virtual object* at O shows the location where the light would be focused if the mirror were not present. Of course, no light is present on the *V*-side of the mirror. A real image I is formed. This arrangement produces a real image only if the magnitude of the object distance is less than the focal length, but in a similar situation a convex lens *always* forms a real image.

The magnification is positive, and thus the image is upright, as shown. Can you predict the outcome if the mirror in Fig. 40-11 were made concave instead of convex? How would the resulting image distance compare in magnitude with the object distance? Would the image be upright or inverted?

Ray Tracing

It is a good idea to check the results of algebraic computations obtained from the mirror equation against a graphical method for locating the image. This method is called *ray tracing.* As suggested in Fig. 40-10, bundles of rays either converge to a real image or diverge from a virtual image. If we can draw these rays as they reflect from the mirror, we can locate the image.

We can simplify this procedure by drawing a few basic rays, the intersection of which serves to locate the image. These rays, which are shown in Fig. 40-12, are selected from an infinite number of possible rays for convenience in locating the image. These rays need not necessarily exist in actuality (part of the mirror might be covered by an opaque screen, for example), but they nevertheless can be used to find the image (which is complete even if some of the rays are blocked). The rays are:

1. *A ray parallel to the axis,* which is reflected to pass through the focal point (in the case of a converging mirror, Fig. 40-12*a*) or to appear to come from the focal point (in the case of a diverging mirror, Fig. 40-12*c*).

2. *A ray passing through the focal point* (converging mirror, Fig. 40-12*a*) or doing so when extended (diverging mirror, Fig. 40-12*c*), which is reflected to be parallel to the axis.

3. *A ray passing through the center of curvature C,* which is reflected back along its original path (Figs. 40-12*b* and 40-12*d*).

4. *A ray striking the vertex of the mirror* (the point v where the axis intersects the mirror), which is reflected at an equal angle on the opposite side of the axis (Figs. 40-12*b* and 40-12*d*).

Any two of these four rays can be used to locate the image, as indicated in Fig. 40-12.

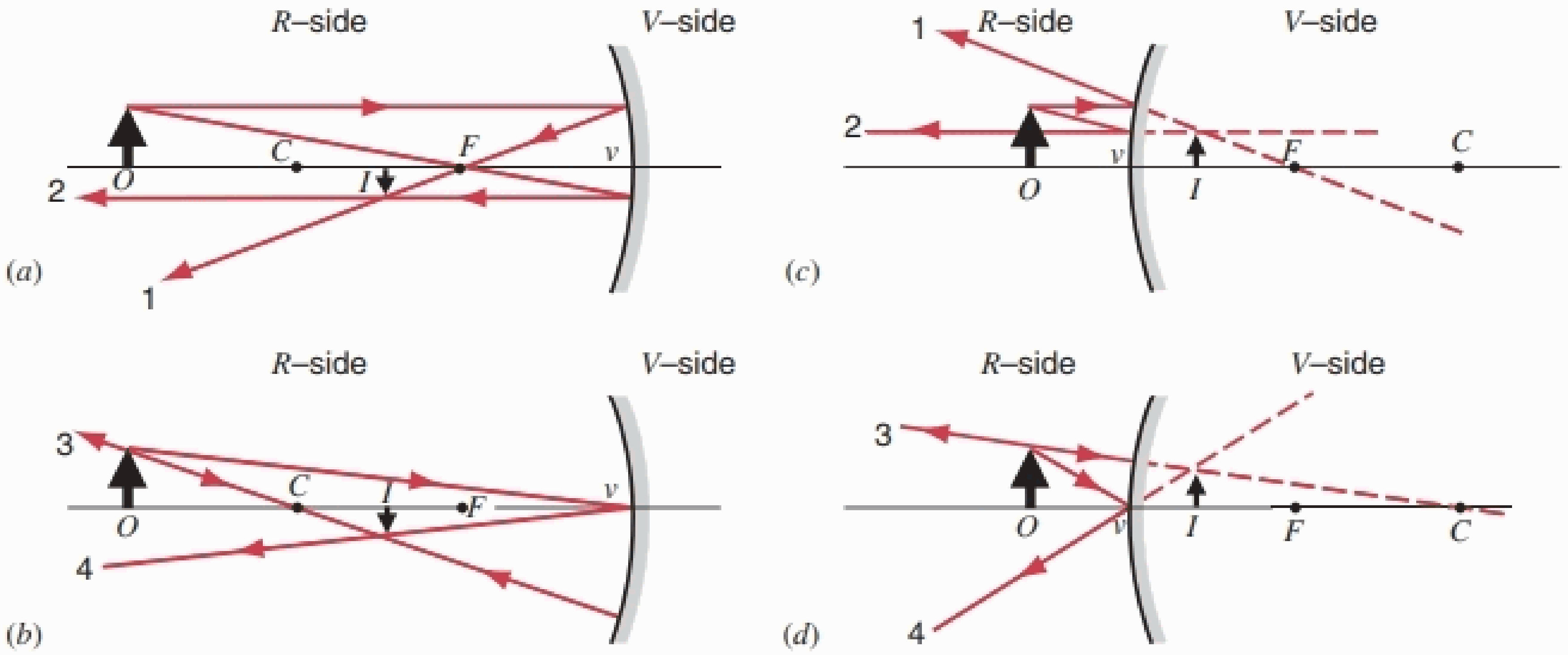


FIGURE 40-12. (*a*, *b*) Four rays that may be used in graphical constructions to locate the image of an object in a concave mirror. Note that the image is real and inverted. (*c*, *d*) Four similar rays drawn in the case of a convex mirror. The image is virtual and upright.

SAMPLE PROBLEM 40-2. In the situation shown in Figs. 40-12*a* and 40-12*b*, suppose $f = 12$ cm and $o = 36$ cm. Find the position of the image and the lateral magnification.

Solution Solving Eq. 40-6 for $1/i$, we obtain

$$\frac{1}{i} = \frac{1}{f} - \frac{1}{o} = \frac{1}{12\text{ cm}} - \frac{1}{36\text{ cm}}$$

or

$$i = 18\text{ cm}.$$

This is consistent with Figs. 40-12*a* and 40-12*b*.

Using Eq. 40-3, which applies to spherical as well as plane mirrors, we have

$$m = -\frac{i}{o} = -\frac{18\text{ cm}}{36\text{ cm}} = -0.50.$$

The image is 1/2 the size of the object and (as indicated by the negative sign) is inverted. These are consistent with Figs. 40-12*a* and 40-12*b*.

SAMPLE PROBLEM 40-3. A convex mirror has a radius of curvature of 22 cm. An object is placed 14 cm from the mirror. Locate and describe the image using (*a*) graphical and (*b*) algebraic methods.

Solution (*a*) Figure 40-13 shows the object and the mirror. Rays 1, 2, and 3 are drawn to locate the image. The image is virtual, upright, and located on the *V*-side of the mirror; the image distance is in magnitude about half the object distance, and the image is about half as tall as the object.

(*b*) According to our sign convention, the radius is negative if the center of curvature is located on the *V*-side of the mirror. Solving Eq. 40-4 for $1/i$, we obtain

$$\frac{1}{i} = \frac{2}{r} - \frac{1}{o} = \frac{2}{-22\text{ cm}} - \frac{1}{+14\text{ cm}},$$

which yields

$$i = -6.2\text{ cm}.$$

This value is consistent with the result of our graphical construction.

The lateral magnification is, from Eq. 40-3,

$$m = -\frac{i}{o} = -\frac{-6.2\text{ cm}}{+14\text{ cm}} = +0.44,$$

which is also consistent with the result obtained graphically. Note that $m > 0$, indicating that the image is upright.

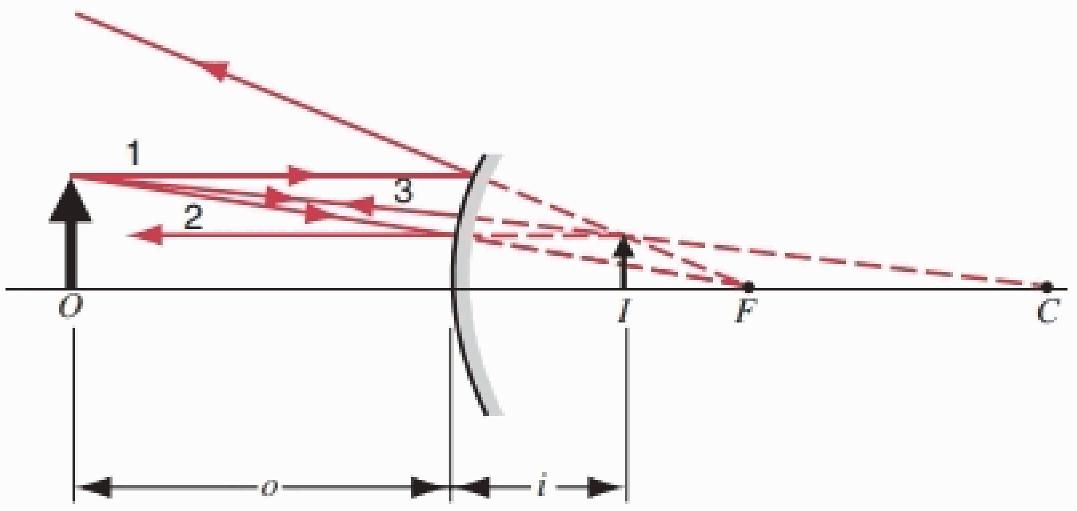


FIGURE 40-13. Sample Problem 40-3.

Derivation of Mirror Equations

Figure 40-14 shows a point object *O* on the axis of a concave spherical mirror whose radius of curvature is *r*. A ray from *O* that makes an arbitrary angle α with the axis intersects the axis at *I* after reflection from the mirror at *a*. A ray that leaves *O* along the axis is reflected back along itself at *v* and also passes through *I*. Thus *I* is the image of *O*; it is a *real* image because light actually passes through *I*. Let us find the location of *I*.

A useful theorem is that the exterior angle of a triangle is equal to the sum of the two opposite interior angles. Applying this to triangles *OaC* and *OaI* in Fig. 40-14 yields

$$\beta = \alpha + \theta \quad \text{and} \quad \gamma = \alpha + 2\theta.$$

Eliminating θ between these equations leads to

$$\alpha + \gamma = 2\beta. \tag{40-7}$$

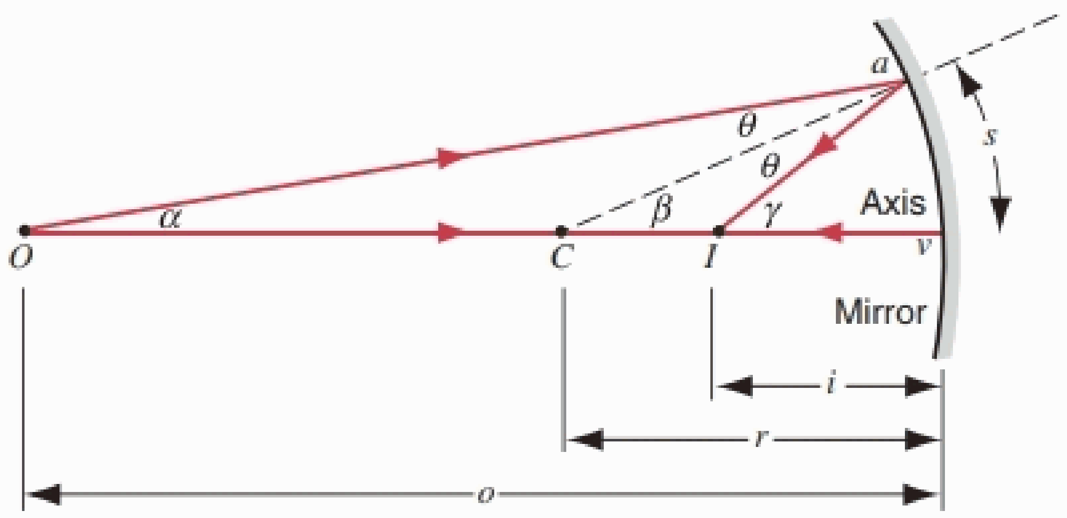


FIGURE 40-14. A point object O forms a real point image I after reflection from a concave mirror.

In radian measure we can write angles α, β, and γ as

$$\alpha \approx \frac{s}{o}, \qquad \beta = \frac{s}{r}, \qquad \gamma \approx \frac{s}{i}, \tag{40-8}$$

where s is the arc length av. Note that only the equation for β is exact, because the center of curvature of arc av is at C and not at O or I. However, the equations for α and for γ are approximately correct if these angles are sufficiently small. *In all that follows we assume that the rays diverging from the object make only a small angle α with the axis of the mirror.* We call such rays, which lie close to the mirror axis, *paraxial* rays. We did not find it necessary to make such an assumption for plane mirrors. Substituting Eqs. 40-8 into Eq. 40-7 and canceling the arc length s yields Eq. 40-4, which is the equation we set out to prove.

Figure 40-15 shows a point object on the axis of a convex mirror. The angles are labeled similarly to those of Fig. 40-14. We can carry out a similar analysis to that given previously, which again yields Eq. 40-4, *provided* we follow the sign convention that i and r are taken to be negative in Fig. 40-15. This derivation is left as an exercise (see Exercise 16).

To derive the equation for lateral magnification (Eq. 40-3), consider Fig. 40-16, which shows a ray (avb) that originates on the tip of the object, reflects from the mirror at point v, and passes through the tip of the image. The law of reflection demands that this ray make equal angles θ with the mirror axis as shown. For the two similar right triangles aOv and bIv we can write

$$\frac{Ib}{Oa} = \frac{vI}{vO}. \tag{40-9}$$

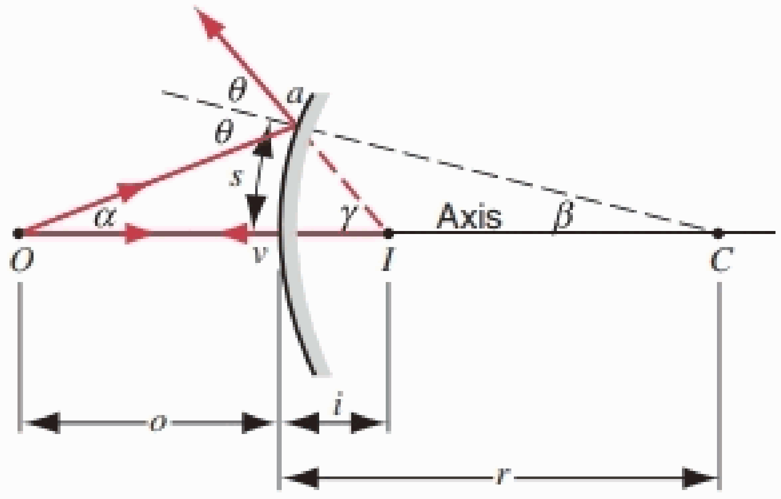


FIGURE 40-15. A point object O forms a virtual point image I after reflection from a convex mirror. Compare with Fig. 40-14.

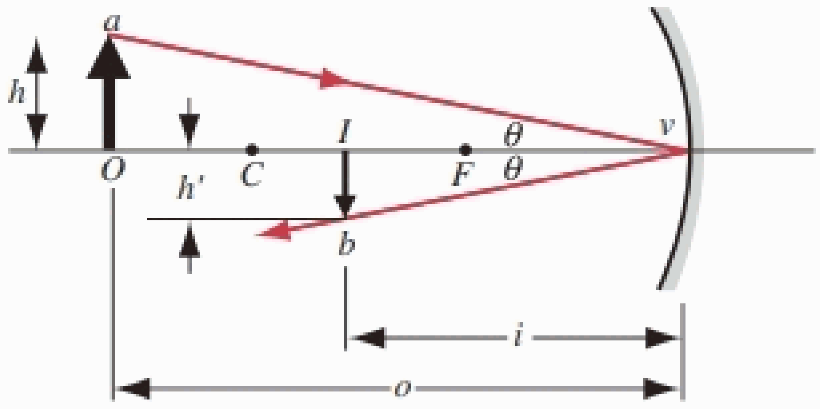


FIGURE 40-16. An object O forms an inverted real image I in a concave mirror.

We take h (the height of the object *above* the axis) to be $+Oa$, and h' (the height of the image *below* the axis) to be $-Ib$. These choices are in agreement with the sign convention discussed for Eq. 40-2. The left side of Eq. 40-9 is then equal to $-h'/h$. With $vI = i$ and $vO = o$, Eq. 40-9 then reduces immediately to $m = -i/o$, which is Eq. 40-3, as we set out to prove.

Significantly, Eq. 40-4 does not contain α (or β, γ, or θ), so that it holds for all rays that strike the mirror provided that they are sufficiently paraxial. In an actual case, the rays can be made as paraxial as one likes by putting a circular diaphragm in front of the mirror, centered about the vertex v; this will impose a certain maximum value of α.

Images in spherical mirrors suffer from several distortions that arise because the assumption of paraxial rays is never completely justified. In general, a point source does not result in a point image; see Exercise 12. Apart from this, distortion arises because the magnification varies somewhat with distance from the mirror axis, Eq. 40-3 being strictly correct only for paraxial rays. Finally, we must always keep in mind that geometrical optics is itself only a special case of physical optics; the effects of diffraction (see Chapter 42) can affect the nature of the image.

40-4 SPHERICAL REFRACTING SURFACES

In Fig. 40-17*a*, light from a point object O falls on a convex spherical refracting surface of radius of curvature r. The surface separates two media; the index of refraction of the medium containing the incident light is n_1, while that of the medium containing the refracted light is n_2. Such a diagram might represent light that is incident on a small region of a glass sphere; note that the real image is formed within the glass (medium 2). Although we do not often encounter images of this type, understanding the images produced by spherical refracting surfaces is essential in the discussion of thin lenses in Section 40-5.

In Fig. 40-17*b*, a concave surface forms a virtual image when $n_1 < n_2$, the light in medium 2 diverging as if it came from the image point I. Figure 40-17*c* shows a surface that is again concave with respect to the incident light, but now $n_1 > n_2$ and a real image is formed.

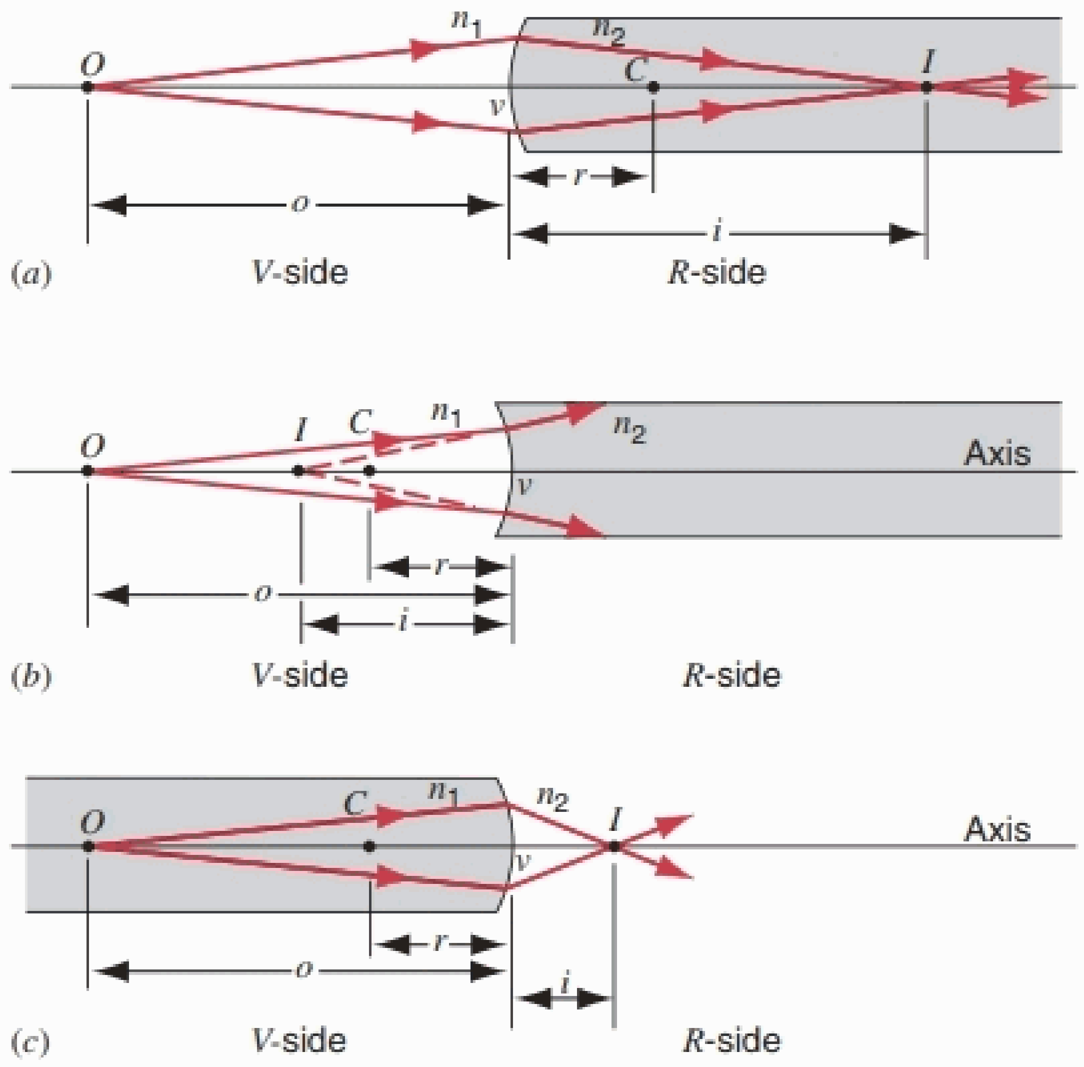


FIGURE 40-17. (*a*) A real image of a point object is formed by refraction at a convex spherical boundary between two media; in this case, $n_2 > n_1$. (*b*) A virtual image is formed by refraction at a concave spherical boundary when $n_2 > n_1$. (*c*) The same as (*b*), except that $n_2 < n_1$. In this case a real image is formed.

As we prove later, the image distance *i* is related to the object distance *o*, radius of curvature *r*, and two indices of refraction according to

$$\frac{n_1}{o} + \frac{n_2}{i} = \frac{n_2 - n_1}{r}. \tag{40-10}$$

This single equation, with appropriate sign conventions, is sufficient to analyze both convex and concave surfaces. The only restriction, as was the case in our discussion of spherical mirrors, is that the rays must be paraxial.

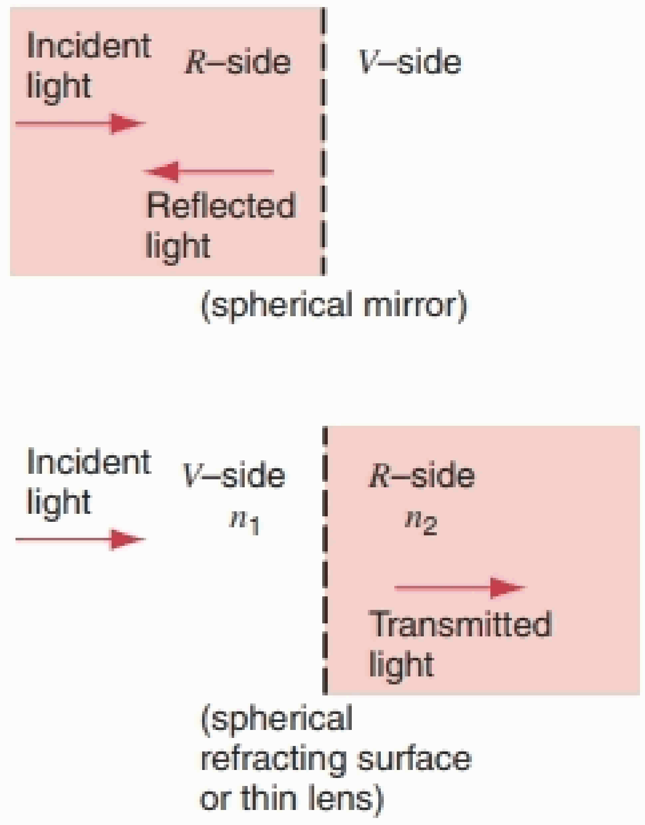


FIGURE 40-18. Real images are formed on the same side as the incident light for mirrors but on the opposite side for refracting surfaces and for lenses.

The sign conventions to be used with Eq. 40-10 are summarized in Fig. 40-18 and compared with those of a spherical mirror. If a real image is to be formed by converging light from the surface, it must appear on the side of the surface opposite to the incident light. This side is called the *R*-side. Virtual images, as shown in Fig. 40-17*b*, are formed on the same side as the incident light, which we call the *V*-side. The radius of curvature is taken as positive if the center of curvature *C* is located on the *R*-side (as in Fig. 40-17*a*) and negative if *C* is located on the *V*-side (as in Figs. 40-17*b* and 40-17*c*). Object distances are positive for real objects (on the *V*-side), whereas image distances are positive for real images (on the *R*-side). The image distance *i* is positive in Fig. 40-17*a* and 40-17*c*, and *i* is negative in Fig. 40-17*b*.

SAMPLE PROBLEM 40-4. Locate the image for the geometry shown in Fig. 40-17*a*, assuming the radius of curvature *r* to be 11 cm, $n_1 = 1.0$, and $n_2 = 1.9$. Let the object be 19 cm to the left of the vertex *v*.

Solution Solving Eq. 40-10 for n_2/i and substituting the given values, we obtain

$$\frac{n_2}{i} = \frac{n_2 - n_1}{r} - \frac{n_1}{o} = \frac{1.9 - 1.0}{+11 \text{ cm}} - \frac{1.0}{+19 \text{ cm}}.$$

Note that *r* is positive because the center of curvature *C* of the surface in Fig. 40-17*a* lies on the *R*-side. If we solve the above equation for *i*, we find

$$i = +65 \text{ cm}.$$

This result is consistent with the sign conventions of Fig. 40-17*a*. The light actually passes through the image point *I* so the image is real, as indicated by the positive sign that we found for *i*. Remember also that n_1 (= 1.0 in this case) always refers to the medium on the side of the surface from which the light comes.

SAMPLE PROBLEM 40-5. A fish is swimming along a horizontal diameter and is 10 cm from the side of a spherical bowl of radius 15 cm (Fig. 40-19). Take the index of refraction of water to be $n_1 = 1.33$ and find the location of the fish according to an

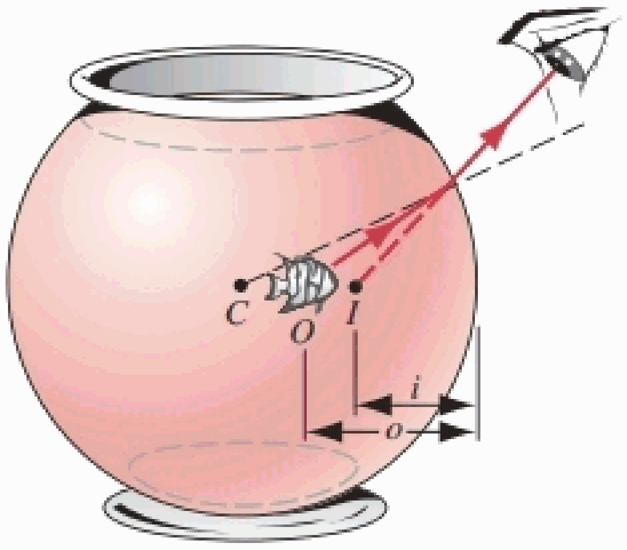


FIGURE 40-19. Sample Problem 40-5. Note that the ray from *O* is bent away from the normal (indicated by the dashed line), in accordance with Snell's law.

observer outside the bowl. Assume the glass bowl is so thin that refraction due to the glass can be ignored.

Solution According to the sign conventions, in the geometry of Fig. 40-19 we take o to be positive because the object is on the V-side of the spherical surface, and we take r to be negative because C is on the V-side. We use Eq. 40-10 with $n_2 = 1$, which we solve to give

$$\frac{n_2}{i} = \frac{n_2 - n_1}{r} - \frac{n_1}{o} = \frac{1 - 1.33}{-15\text{ cm}} - \frac{1.33}{10\text{ cm}}.$$

Solving, we find

$$i = -9.0\text{ cm}.$$

That is, the fish appears closer to the side of the bowl than it really is.

Derivation of Refracting Surface Formula

Figure 40-20 shows a point source O near a spherical refracting surface of radius r. A ray from O strikes the surface at v at normal incidence and passes undeviated into medium 2 through the center of curvature C. This ray establishes a convenient axis for our calculation. A second ray, which makes a small but arbitrary angle α with the axis and strikes the refracting surface at a, is refracted according to

$$n_1 \sin\theta_1 = n_2 \sin\theta_2.$$

The refracted ray intersects the first ray at I, thereby locating the image of O.

As in the derivation of the mirror equation, we use the theorem that the exterior angle of a triangle is equal to the sum of the two opposite interior angles. Applying this to triangles COa and ICa yields

$$\theta_1 = \alpha + \beta \quad\text{and}\quad \beta = \theta_2 + \gamma. \qquad (40\text{-}11)$$

As we did in Section 40-3, we assume all rays are paraxial, so that all angles (β, γ, θ_1, θ_2) are small, and the sine of each angle can be replaced by the angle itself. This permits us to write the law of refraction as

$$n_1\theta_1 \approx n_2\theta_2. \qquad (40\text{-}12)$$

Combining Eqs. 40-11 and 40-12 leads, after rearrangement, to

$$n_1\alpha + n_2\gamma = (n_2 - n_1)\beta. \qquad (40\text{-}13)$$

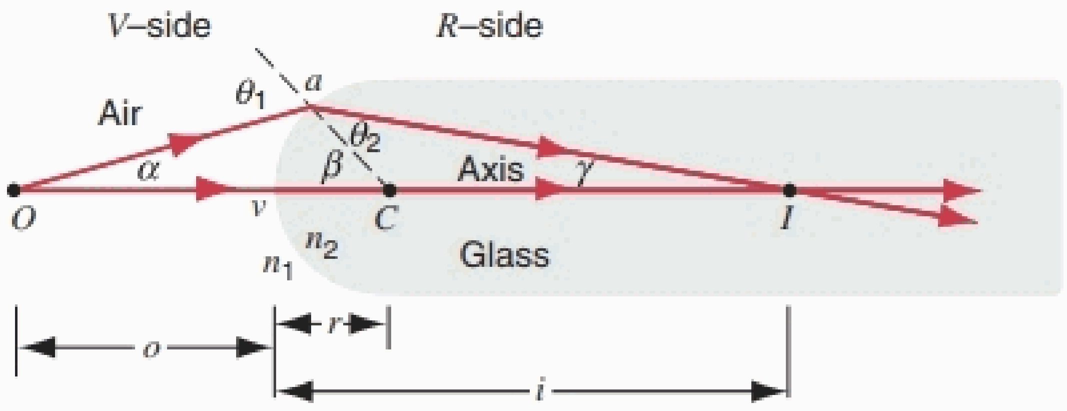


FIGURE 40-20. A point object O forms a real point image I after refraction at a spherical convex surface between two media.

In radian measure the angles α, β, and γ in Fig. 40-20 are

$$\alpha \approx \frac{s}{o}, \quad \beta = \frac{s}{r}, \quad\text{and}\quad \gamma \approx \frac{s}{i}, \qquad (40\text{-}14)$$

where s is the arc length av. Only the second of these equations is exact. The other two are approximate because I and O are *not* the centers of circles of which av is an arc. However, for paraxial rays (α small enough) the inaccuracies in Eqs. 40-14 can be made as small as desired.

Substituting Eqs. 40-14 into Eq. 40-13 leads directly to Eq. 40-10.

40-5 THIN LENSES

There are many common examples of the refraction of light by a lens. The lenses in our eyes focus light on the retina, and the corrective lenses of eyeglasses or contact lenses compensate for deficiencies in our vision. The multi-element lens of a camera focuses light on the film. In this section we consider the properties of such lenses.

In most refraction situations there is more than one refracting surface. This is true even for a contact lens, where the light passes first from air into glass and then from glass into the eye. We consider here only the special case of a *thin lens;* that is, the thickness of the lens is small compared with the object distance o, the image distance i, or the radii of curvature r_1 and r_2 of either of the two refracting surfaces. For such a lens—as we shall prove later in this section—these quantities are related by

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f}, \qquad (40\text{-}15)$$

in which the focal length f of the lens is given by

$$\frac{1}{f} = (n - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right). \qquad (40\text{-}16)$$

Equations 40-15 and 40-16 are approximations that hold only for thin lenses and paraxial rays. Note that Eq. 40-15 is the same equation that we used for spherical mirrors. Equation 40-16 is often called the *lens maker's equation;* it relates the focal length of the lens to the index of refraction n of the lens material and the radii of curvature of the two surfaces.

In Eq. 40-16, r_1 is the radius of curvature of the lens surface on which the light first falls and r_2 is that of the second surface. Equation 40-16 is used in cases in which a lens of index of refraction n is immersed in air. If the lens is immersed in a medium for which the index of refraction is not unity, Eq. 40-16 still holds if we replace n in that formula by $n_{\text{lens}}/n_{\text{medium}}$.

The lateral magnification of a thin lens is given by the same formula as that of a spherical mirror,

$$m = -\frac{i}{o}. \qquad (40\text{-}17)$$

Later in this section we derive this result.

Sign Conventions

The sign conventions for o, i, r_1, and r_1 are similar to those for spherical mirrors and refracting surfaces; see Fig. 40-18. Figure 40-21 illustrates these sign conventions. As before, we have an R-side and a V-side.

1. The radii of curvature r_1 (referring to the first surface struck by the light) and r_2 (referring to the second surface struck by the light) are positive if the corresponding centers of curvature are on the R-side. The radii are negative if the corresponding centers of curvature are on the V-side. In Fig. 40-21*a*, the center of curvature C_1 is on the R-side, so r_1 is positive, and C_2 is on the V-side, so r_2 is negative. Inspection of Eq. 40-16 shows that, when $r_1 > 0$ and $r_2 < 0$, the focal length f is always positive. Such a lens is called a *converging* lens; a lens that is thicker at the center than at the edges, when immersed in a medium of index of refraction lower than that of the lens, is always a converging lens.

In Fig. 40-21*b*, C_1 is on the V-side, and C_2 is on the R-side. Hence r_1 is negative and r_2 is positive. In this case, Eq. 40-16 shows that f is always negative. Such a lens is called a *diverging* lens; a lens that is thinner at the center than at the edges, when immersed in a medium of lower index of refraction, is always a diverging lens.

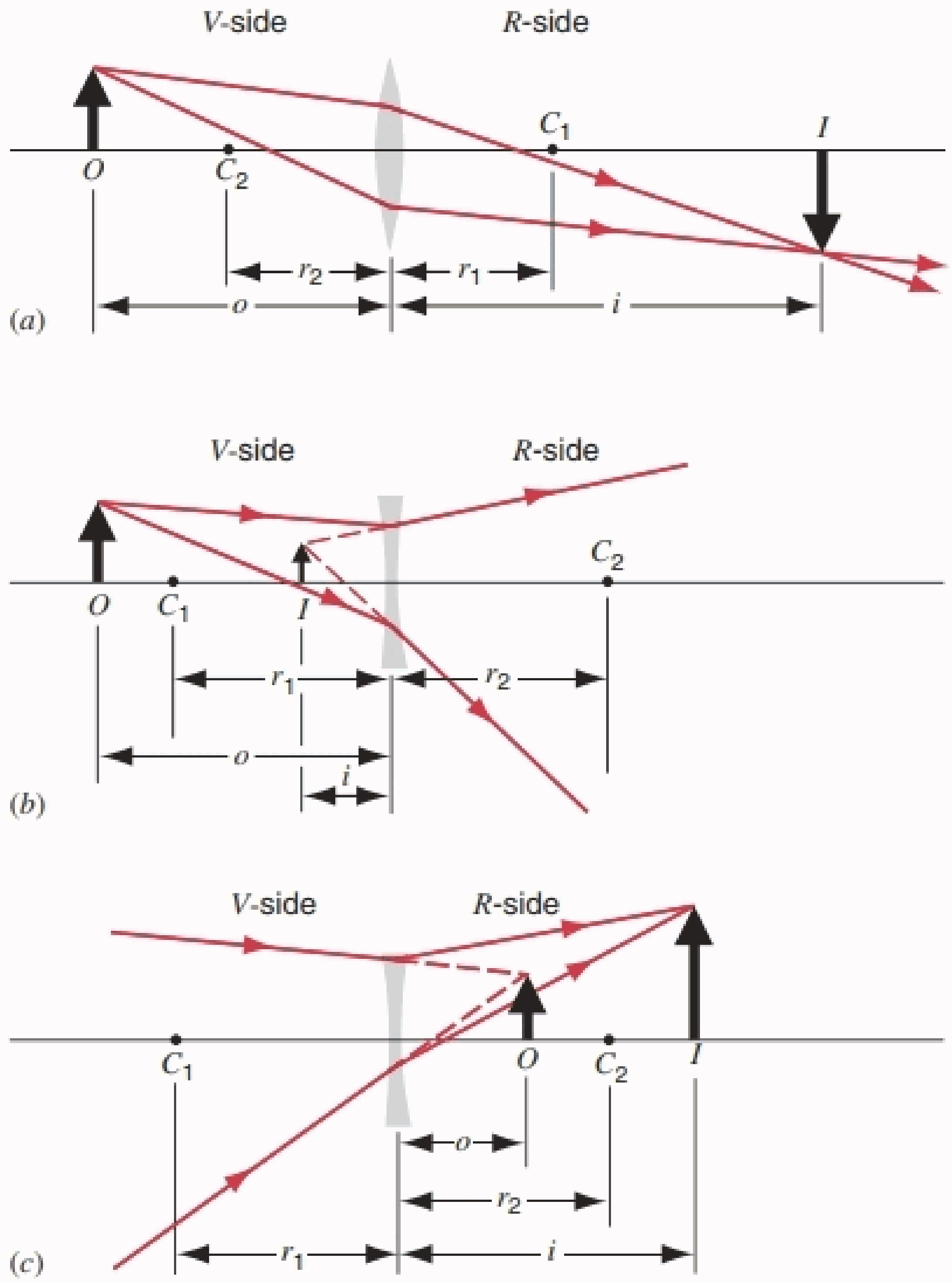


FIGURE 40-21. (*a*) A real, inverted image is formed by a converging lens. Such a lens has a positive focal length and is thicker at the center than at the edges. (*b*) A virtual, erect image is formed by a diverging lens. Such a lens has a negative focal length and is thinner at the center than at the edges. (*c*) Converging light gives a virtual object at O. A real, erect image is formed at I by this diverging lens.

2. The object distance o is positive if the object is real and lies on the V-side of the lens, as in both Figs. 40-21*a* and 40-21*b*. Light from a real object is diverging when it strikes the lens. It is also possible to have converging light strike the lens, as in Fig. 40-21*c*. In this case, if the lens were not present, the converging light would form an image at O on the R-side of the lens; we take this image as a virtual object, and we take o as negative in this case.

3. The image distance i is positive if the (real) image lies on the R-side of the lens, as in Figs. 40-21*a* and 40-21*c*, whereas i is negative if the (virtual) image lies on the V-side of the lens, as in Fig. 40-21*b*.

4. According to Eq. 40-17, the magnification is negative when both i and o are positive, as in Fig. 40-21*a*, corresponding to an *inverted* image. In the case of an erect image, as in Figs. 40-21*b* and 40-21*c*, the magnification is positive, because o and i have opposite signs. In the case shown in Fig. 40-21*b*, o is positive and i is negative, whereas in Fig. 40-21*c*, o is negative and i is positive.

In contrast with a spherical mirror or a spherical refracting surface, a lens has *two* focal points. In a thin lens, the two focal points are located at equal distances f from the lens on either side of the lens.

When a point object is located at the *first focal point* F_1, parallel light emerges from the lens, as shown in Fig. 40-22*a*. In the case of a diverging lens (Fig. 40-22*b*), the point object is a virtual object. Converging light, which would have been focused at F_1 if the lens were not there, is defocused into parallel light by the diverging lens. The *second focal point* F_2 is the point where parallel light is (or appears to be) focused by the lens, as shown in Fig. 40-23. Note

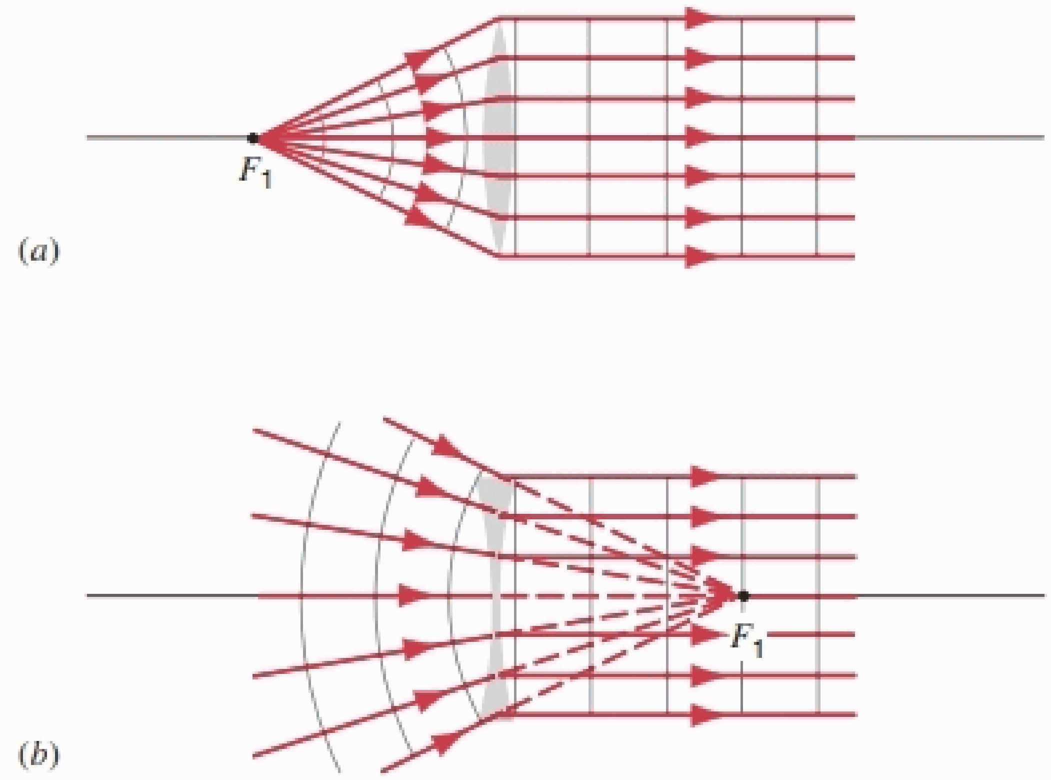


FIGURE 40-22. (*a*) When a point object is at the first focal point F_1 of a converging lens, parallel light emerges from the lens. (*b*) In the case of a diverging lens, a virtual point object gives parallel light.

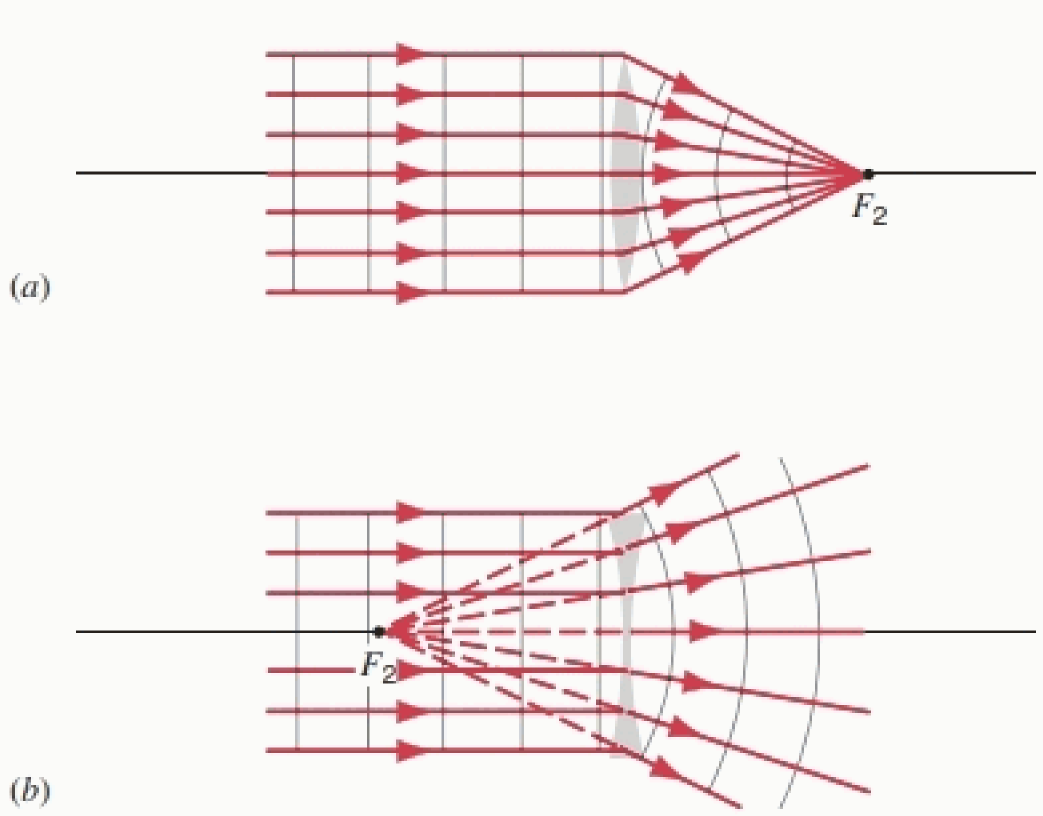


FIGURE 40-23. (*a*) When parallel light is incident on a converging lens, the light is brought to a focus at the second focal point F_2. (*b*) When parallel light is incident on a diverging lens, it appears to emerge from the second focal point.

from comparing Figs. 40-22 and 40-23 that the locations of the first and second focal points in a converging lens are opposite to those in a diverging lens.

In Figs. 40-22 and 40-23, all rays contain the same number of wavelengths; that is, they have the same *optical path length* (see Eq. 39-16). Note how the different geometrical lengths of the paths of the rays through the lens (where the speed of light is smaller than it is in air) changes the spherical wavefronts into planes or the plane wavefronts into spherical ones.

Ray Tracing

As was the case with spherical mirrors, it is helpful to locate the image formed by a thin lens using a graphical method with a few basic rays. Figure 40-24 shows three rays that can be used as applied to a real image formed by a converging lens (Fig. 40-24*a*), a virtual image formed by a converging lens (Fig. 40-24*b*), and a virtual image formed by a diverging lens (Fig. 40-24*c*):

1. A ray (ray 1 in Fig. 40-24) passing through (or, when extended, appearing to pass through) the first focal point F_1 emerges from the lens parallel to the axis.

2. A ray (ray 2 in Fig. 40-24) parallel to the axis passes through (or, when extended, appears to pass through) the second focal point F_2.

3. A ray (ray 3 in Fig. 40-24) falling on the lens at its center passes through the lens undeflected, because near its center the lens behaves like a flat piece of glass with parallel sides, which does not change the direction of the ray.

Any two of these rays can be used to locate the image; the third is available as a check. Note from Fig. 40-24 that for all three rays, we consider the refraction to take place in a plane at the location of the lens. This can be done only for a thin lens.

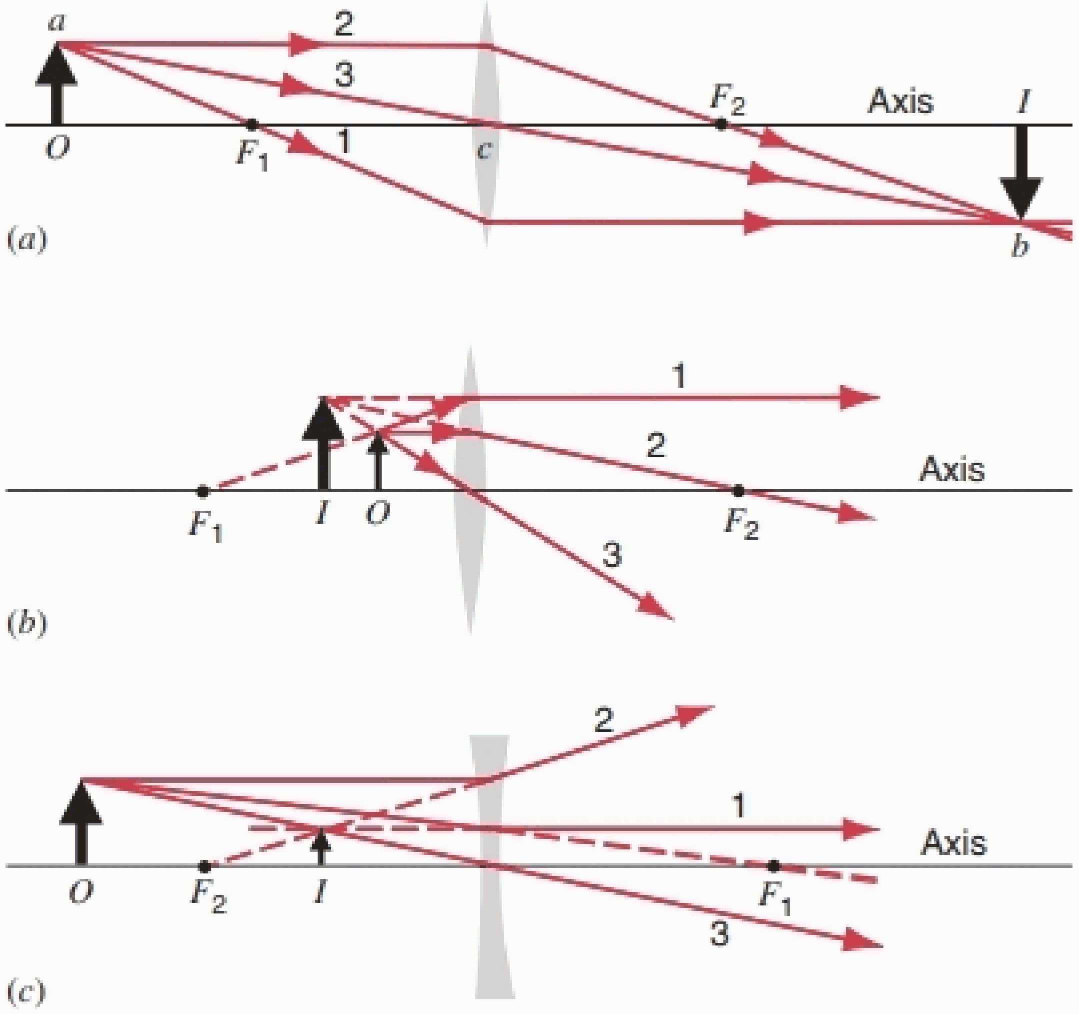


FIGURE 40-24. Three rays (1, 2, 3) that can be used to locate the image formed by (*a*) a converging lens (real image), (*b*) a converging lens (virtual image), and (*c*) a diverging lens (virtual image).

Derivation of the Thin-Lens Formulas

Our plan is to consider each lens surface separately, using the image formed by the first surface as an object for the second.

Figure 40-25*a* shows such a thick glass "lens" of length L whose surfaces are ground to radii r_1 and r_2. A point object O is placed near the left surface as shown. A ray leaving O along the axis is not deflected on entering or leaving the lens.

A second ray leaving O at an angle α with the axis strikes the surface at point a, is refracted, and strikes the second surface at point b. The ray is again refracted and crosses the axis at I, which, being the intersection of two rays from O, is the image of point O, formed after refraction at two surfaces.

Figure 40-25*b* shows the first surface, which forms a virtual image of O at I'. To locate I', we use Eq. 40-10, with $n_1 = 1$ and $n_2 = n$:

$$\frac{1}{o} + \frac{n}{i'} = \frac{n-1}{r_1},$$

or, taking into account that i' is negative,

$$\frac{1}{o} - \frac{n}{|i'|} = \frac{n-1}{r_1}. \qquad \text{(40-18)}$$

Figure 40-25*c* shows the second surface. An observer at point *b*, unaware of the existence of the first surface, would think that the light striking that point originated at point I' in Fig. 40-25*b* and that the region to the left of the surface was filled with glass. Thus the (virtual) image I' formed by the first surface serves as a real object O' for the second

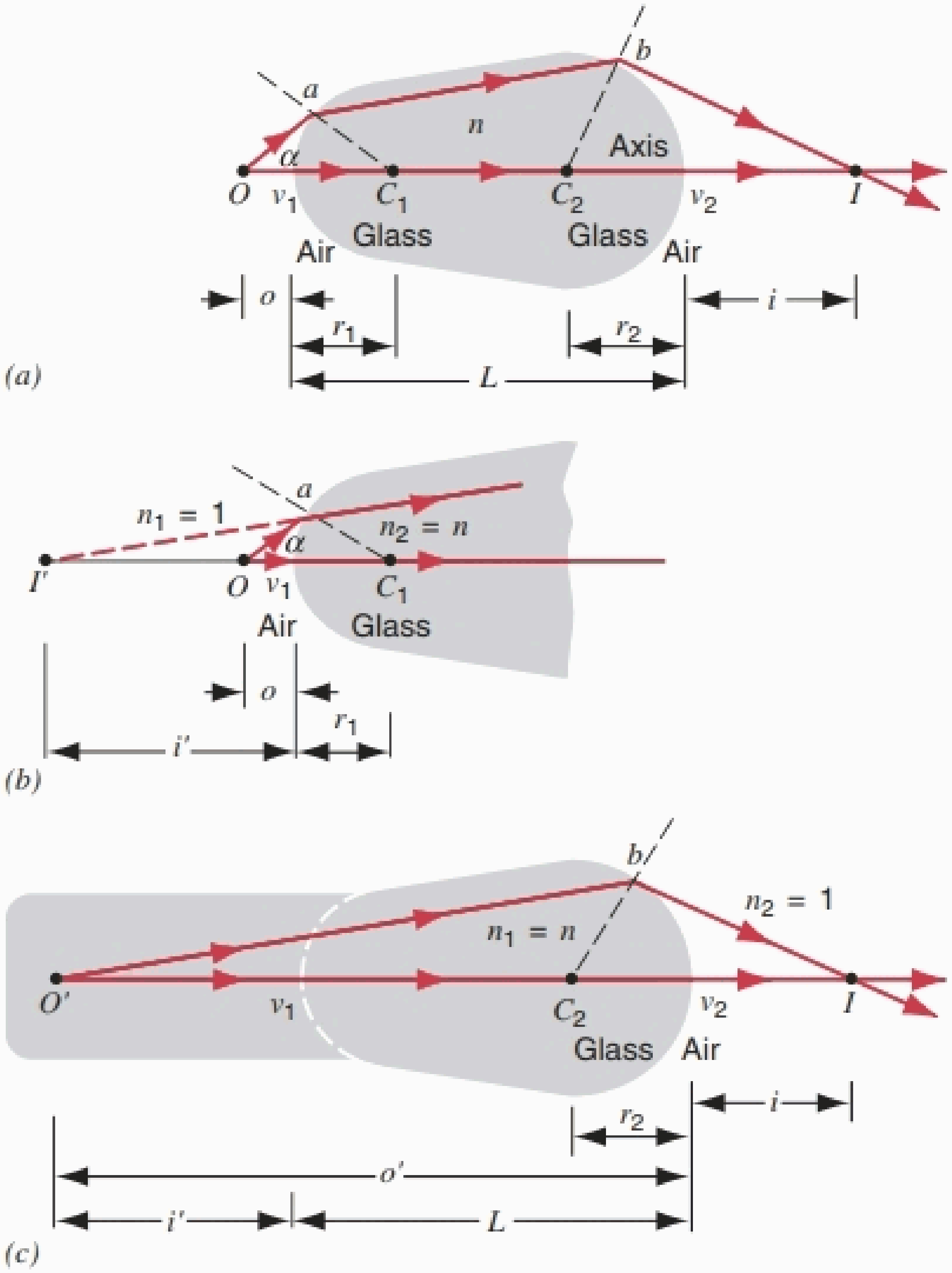


FIGURE 40-25. (*a*) Two rays from *O* form a real image at *I* after refraction at two spherical surfaces, the first surface being converging and the second diverging. (*b*) The first surface and (*c*) the second surface, shown separately. The vertical scale has been greatly exaggerated for clarity.

surface. The distance of this object from the second surface is

$$o' = |i'| + L. \tag{40-19}$$

In applying Eq. 40-10 to the second surface, we insert $n_1 = n$ and $n_2 = 1$ because the object behaves as if it were imbedded in glass. If we use Eq. 40-19, Eq. 40-10 becomes

$$\frac{n}{|i'| + L} + \frac{1}{i} = \frac{1 - n}{r_2}. \tag{40-20}$$

Let us now assume that the thickness L of the "lens" in Fig. 40-25*a* is so small that we can neglect it in comparison with other linear quantities in this figure (such as o, i, o', i', r_1, and r_2). In all that follows we make this *thin-lens approximation.* Putting $L = 0$ in Eq. 40-20 leads to

$$\frac{n}{|i'|} + \frac{1}{i} = -\frac{n - 1}{r_2}. \tag{40-21}$$

Adding Eqs. 40-18 and 40-21 gives

$$\frac{1}{o} + \frac{1}{i} = (n - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right). \tag{40-22}$$

Defining the right side of Eq. 40-22 to be $1/f$ leads directly to Eqs. 40-15 and 40-16, completing the derivation.

To derive Eq. 40-17 for the lateral magnification, we refer to Fig. 40-24*a*. Right triangles *acO* and *bcI* are similar, because angles *acO* and *bcI* are equal. For the corresponding sides of the similar triangles, we obtain

$$\frac{bI}{aO} = \frac{cI}{cO}. \tag{40-23}$$

The right side of this expression is simply i/o, while the left side is $-m$, the negative sign indicating that the image is inverted. With these substitutions, Eq. 40-23 reduces directly to Eq. 40-17.

SAMPLE PROBLEM 40-6. The lenses of Fig. 40-21 have radii of curvature of magnitude 42 cm and are made of glass with $n = 1.65$. Compute their focal lengths.

Solution Since C_1 lies on the *R*-side of the lens in Fig. 40-21*a*, r_1 is positive ($= +42$ cm). Since C_2 lies on the *V*-side, r_2 is negative ($= -42$ cm). Substituting in Eq. 40-16 yields

$$\frac{1}{f} = (n - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = (1.65 - 1)\left(\frac{1}{+42\text{ cm}} - \frac{1}{-42\text{ cm}}\right)$$

or

$$f = +32\text{ cm}.$$

A positive focal length indicates that, in agreement with what we have said, parallel incident light converges after refraction to form a real focus.

In Figs. 40-21*b* and 40-21*c*, C_1 lies on the *V*-side of the lens so that r_1 is negative ($= -42$ cm). Since r_2 is positive ($= +42$ cm), Eq. 40-16 yields

$$\frac{1}{f} = (n - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = (1.65 - 1)\left(\frac{1}{-42\text{ cm}} - \frac{1}{+42\text{ cm}}\right)$$

or

$$f = -32\text{ cm}.$$

SAMPLE PROBLEM 40-7. An object is 38 cm in front of a diverging lens of focal length -24 cm. Find the location and lateral magnification of the image.

Solution Using Eq. 40-15, we have

$$\frac{1}{+38\text{ cm}} + \frac{1}{i} = \frac{1}{-24\text{ cm}}$$

or

$$i = -15\text{ cm}.$$

The magnification is

$$m = -\frac{i}{o} = -\frac{-15\text{ cm}}{+38\text{ cm}} = +0.39.$$

Figure 40-24*c* illustrates the graphical solution to this problem. You can see that the location, size, and orientation of the image are consistent with the results of the calculation.

Two-Lens Systems

A single mirror or lens is rarely a useful optical device. In such instruments as binoculars, telescopes, microscopes, and cameras, images are formed by a combination of several lenses or mirrors. In this section we consider the images formed by systems containing two lenses.

To analyze the formation of images by a two-lens system, we consider the lenses one at a time, as if the other were not present, taking the *image* formed by one lens as the *object* for the next. We apply the previously derived formulas for the thin lens (Eqs. 40-15 and 40-17), taking careful account of the sign conventions in each case. In particular, note the following:

> *When* diverging *light from the image formed by one lens strikes the next lens, we treat that image as a* real object *for the second lens. When* converging *light from the image formed by one lens strikes the next element, we treat that image as a* virtual object *for the second lens.*

The total lateral magnification m_t of the two-lens system is equal to the product of the magnifications of each of the individual lenses:

$$m_t = mm', \tag{40-24}$$

where m and m' represent the lateral magnifications of the two individual lenses.

It is straightforward to extend these methods to systems of one lens and one mirror, or to systems with more than two components.

SAMPLE PROBLEM 40-8. Two identical converging lenses of focal lengths $f = f' = +15$ cm are separated by a distance d of 6 cm, as shown in Fig. 40-26. A luminous source is placed a distance of $o = 10$ cm from the first lens. Locate the final image.

Solution We begin by locating the image using a ray diagram, as shown in Fig. 40-26. Rays 2 and 3 from the object O are refracted as shown by the first lens; extended backward they show the location of the (virtual) image I produced by the first lens. This image then acts as the object O' of the second lens, and rays $2'$ and $3'$ give the position of the final image I', which is inverted and real.

Applying Eq. 40-15, we can find the position of the first image:

$$\frac{1}{+10\text{ cm}} + \frac{1}{i} = \frac{1}{+15\text{ cm}}$$

or

$$i = -30\text{ cm}.$$

That is, the image is virtual and formed 30 cm from the lens on its V-side. Treating this image as the object O' for the second lens, the object distance o' is $|i| + d = 30\text{ cm} + 6\text{ cm} = 36$ cm. Note that, although I is a *virtual* image for the first lens, O' is a *real* object for the second lens, because *diverging* light leaves this object and strikes the second lens.

Applying Eq. 40-15 once again, we have

$$\frac{1}{+36\text{ cm}} + \frac{1}{i'} = \frac{1}{+15\text{ cm}}$$

or

$$i' = +26\text{ cm},$$

corresponding to a real image on the R-side of the second lens, as shown in Fig. 40-26.

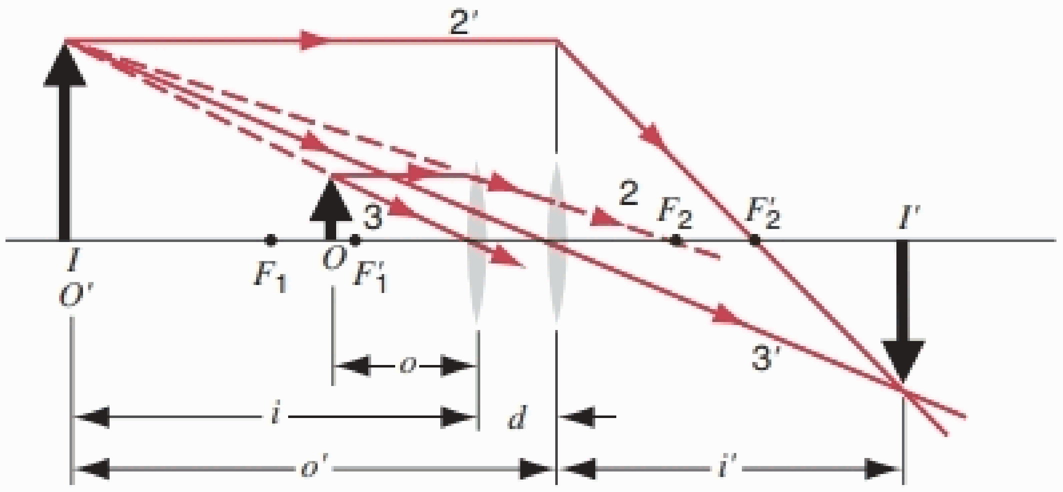


FIGURE 40-26. Sample Problem 40-8.

SAMPLE PROBLEM 40-9. The optical system shown in Fig. 40-27 consists of two lenses, of focal lengths $f = +12$ cm and $f' = -32$ cm, separated by a distance of $d = 22$ cm. A luminous object is placed 18 cm from the first lens. Locate the final image produced by this system.

Solution A graphical construction using a ray diagram is given in Fig. 40-27. The real image I of the first lens would form to the right of the second lens. Because light forming this image is converging as it strikes the second lens, we treat it as a virtual object O' for the second lens.

For the first lens, Eq. 40-15 gives

$$\frac{1}{+18\text{ cm}} + \frac{1}{i} = \frac{1}{+12\text{ cm}}$$

or

$$i = +36\text{ cm}.$$

The real image would be formed 36 cm from the first lens, as shown. The distance o' from the virtual object O' to the second lens has magnitude $i - d$ or $36\text{ cm} - 22\text{ cm} = 14$ cm. Because O' is a virtual object, we take the distance o' to be negative. Now Eq. 40-15 gives, for the second lens,

$$\frac{1}{-14\text{ cm}} + \frac{1}{i'} = \frac{1}{-32\text{ cm}}$$

or

$$i' = +25\text{ cm}.$$

The real image I' forms on the R-side of the second lens.

SAMPLE PROBLEM 40-10. The object in Fig. 40-27 has a height h of 2.4 cm. Find the height of the image.

Solution We seek the lateral magnification of the compound system. Once again, we treat the compound system as two separate systems, and the total lateral magnification m_t of the combined

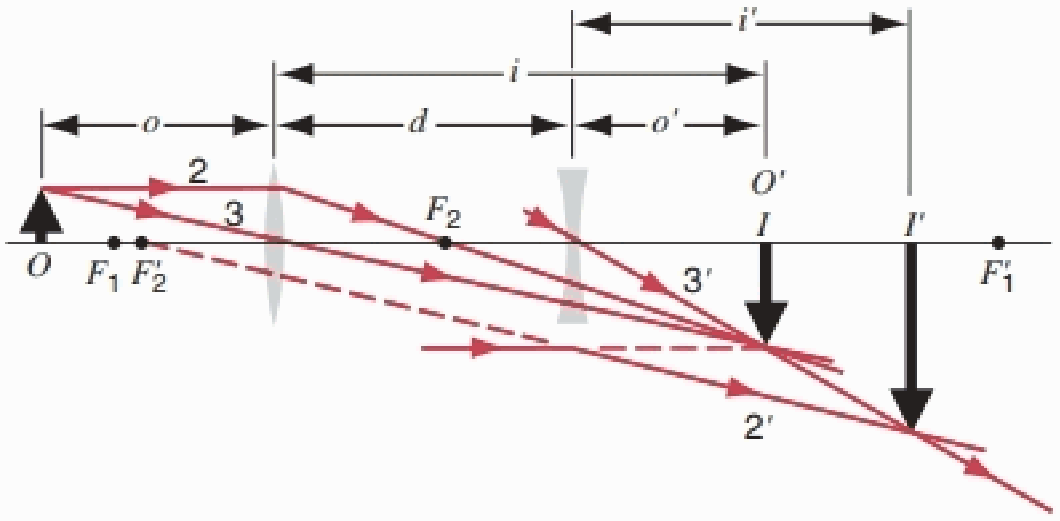


FIGURE 40-27. Sample Problems 40-9, 40-10.

system is the product of the lateral magnifications m and m' of the individual lenses:

$$m_t = mm' = \left(-\frac{i}{o}\right)\left(-\frac{i'}{o'}\right) = \left(-\frac{+36\text{ cm}}{+18\text{ cm}}\right)\left(-\frac{+25\text{ cm}}{-14\text{ cm}}\right)$$

$$= -3.57,$$

where we have used the values of the object and image distances found in Sample Problem 40-9. The height h_f of the final image is

$$h_f = m_t h = (-3.57)(2.4\text{ cm}) = -8.6\text{ cm}.$$

The negative value reminds us that the final imge is inverted with respect to the original object.

40-6 OPTICAL INSTRUMENTS

The human eye is a remarkably effective organ, but its range can be extended in many ways by optical instruments such as eyeglasses or contact lenses, simple magnifiers, motion picture projectors, cameras (including TV cameras), microscopes, and telescopes. In many cases these devices extend the scope of our vision beyond the visible range; satellite-borne infrared cameras and x-ray microscopes are examples.

In almost all cases of modern sophisticated optical instruments, the mirror and thin lens formulas hold only as approximations. In typical laboratory microscopes the lens can by no means be considered "thin." In most optical instruments lenses are compound; that is, they are made of several components.

In this section we consider optical devices that are designed to produce an enlarged image; we want something to *appear* larger than it appears to the unaided eye. The lateral magnification is an incomplete measure of the apparent size of an image produced by an optical system. An optical system might produce an enlarged image ($|m| > 1$), but may place that image so much farther from us than the object that it would actually appear to the observer to be smaller than the object. Even though the lateral magnification may be greater than unity, and thus the image size greater than the object size, the net result is not what the observer would call a "magnified" image.

The Simple Magnifier

Figure 40-28*a* represents the formation of an image by a human eye. The size of the image on the retina is determined by the angle θ subtended at the eye by the object. For small objects located at relatively large distances from the eye, the angle θ can be approximated as

$$\theta \approx \frac{h}{d}, \tag{40-25}$$

where h is the size of the object and d is its distance from the eye.

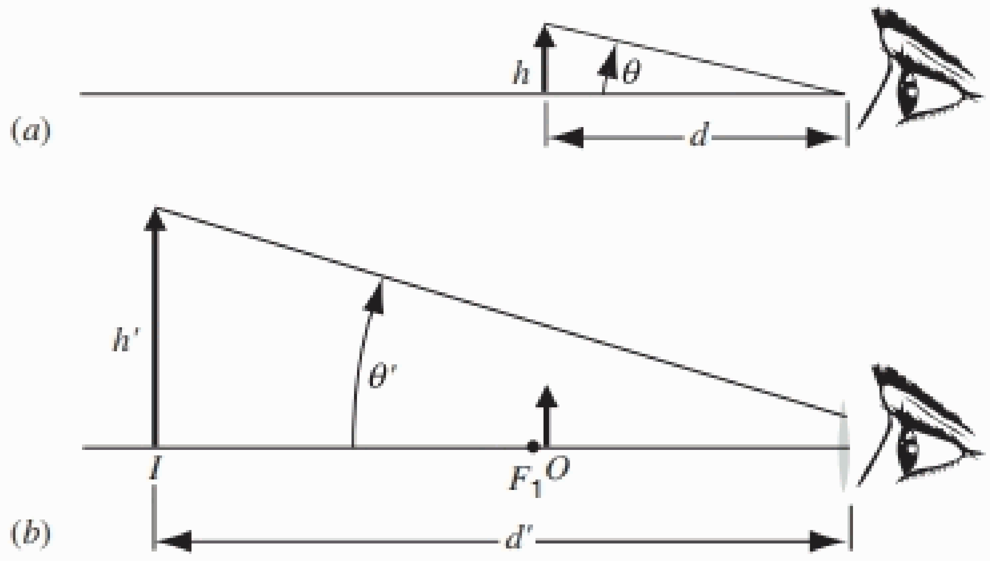


FIGURE 40-28. (*a*) An object of height h at a distance d from the eye subtends an angle θ. (*b*) When the object is viewed through a lens used as a simple magnifier, the image I of height h' is at a distance d' and subtends an angle θ' at the eye.

In Fig. 40-28*b*, the observer is viewing the object through a lens that forms an image of lateral size h' at a distance d' from the eye. The apparent angular size of the image to the observer is, again for small angles,

$$\theta' \approx \frac{h'}{d'}. \tag{40-26}$$

The image viewed through the lens will *appear* larger than the original object to the observer if it subtends a larger solid angle than the object subtends. It is therefore not the lateral magnification m ($= h'/h$) that is important in measuring the apparent size of the image; it is the *angular magnification* m_θ, defined as

$$m_\theta = \frac{\theta'}{\theta}. \tag{40-27}$$

In effect, m_θ is the ratio of the size of the two images on the retina, one with the lens and one without.

The normal human eye can focus a sharp image of an object on the retina if the object O is located anywhere from infinity (the stars, say) to a certain point called the *near point* P_n, which we take to be about 25 cm from the eye. If you view an object closer than the near point, the perceived retinal image becomes fuzzy. The location of the near point normally varies with age. We have all heard stories about people who claim not to need glasses but who read their newspapers at arm's length; their near points are receding! Find your own near point by moving this page closer to your eyes, considered separately, until you reach a position at which the image begins to become indistinct.

We take as our basis for comparison the angular size that an object would appear to have if it were placed at the near point. Thus

$$\theta = \frac{h}{25\text{ cm}}. \tag{40-28}$$

If we place the object so that it is just inside the first focal point of a converging lens, as in Fig. 40-28*b*, a virtual image is formed far away from the lens. The lateral magnification m has magnitude i/o, and the distance d' to the image is i. The lateral size of the image is, taking magnitudes of all quantities,

$$h' = mh = \frac{i}{o}h \tag{40-29}$$

and the angular size is

$$\theta' = \frac{h'}{d'} = \frac{(i/o)h}{i} = \frac{h}{o} \approx \frac{h}{f}, \tag{40-30}$$

where the last step can be taken because we asumed the object to be placed close to the focal point. The angular magnification is

$$m_\theta = \frac{\theta'}{\theta} = \frac{h/f}{h/25\text{ cm}}$$

or

$$m_\theta = \frac{25\text{ cm}}{f}. \tag{40-31}$$

Equation 40-31 gives the angular magnification of the *simple magnifier,* which uses only one lens. The ordinary "magnifying glass," used by stamp collectors and actors portraying Sherlock Holmes, is in reality a simple magnifier. To obtain large angular magnification, we want f to be as small as possible. In practice, an angular magnification of about 10 is the best we can do before lens aberrations begin to distort the image. More sophisticated magnifiers, such as the compound microscope discussed next, can have appreciably greater angular magnifications.

Compound Microscope

Figure 40-29 shows a thin-lens version of a compound microscope, used for viewing small objects that are very close to the objective lens of the instrument. The object O, of height h, is placed just outside the first focal point F_1 of the objective lens, whose focal length is f_{ob}. A real, inverted image I of height h' is formed by the objective, with the lateral magnification given by Eq. 40-17, or

$$m = \frac{h'}{h} = \frac{-s\tan\theta}{f_{\text{ob}}\tan\theta} = -\frac{s}{f_{\text{ob}}}. \tag{40-32}$$

As usual, the negative sign indicates an inverted image.

The distance s (called the *tube length*) is chosen so that the image I falls close to the first focal point F_1' of the eyepiece, which then acts as a simple magnifier as described previously. Parallel rays enter the eye, and a final image I' forms almost at infinity. The final magnification M is the product of the linear magnification m for the objective lens (Eq. 40-32) and the angular magnification m_θ of the eyepiece (Eq. 40-31), or

$$M = mm_\theta = -\frac{s}{f_{\text{ob}}}\frac{25\text{ cm}}{f_{\text{ey}}}. \tag{40-33}$$

Refracting Telescope

Like microscopes, telescopes come in a large variety of forms. The form we describe here is the simple refracting telescope consisting of an objective lens and an eyepiece, both represented in Fig. 40-30 by thin lenses. In practice, just as in microscopes, each lens may be a compound lens system.

At first glance it may seem that the lens arrangements for telescope and for microscopes are similar. However, telescopes are designed to view large objects, such as galaxies, stars, and planets, at large distances, whereas microscopes are designed for the opposite purpose. Note also that in Fig. 40-30 the second focal point of the objective F_2 coincides with the first focal point of the eyepiece F_1', but in Fig. 40-29 these points are separated by the tube length s.

In Fig. 40-30, parallel rays from a distant object strike the objective lens, making an angle θ_{ob} with the telescope axis and forming a real, inverted image at the common focal point F_2, F_1'. This image acts as an object for the eyepiece and a (still inverted) virtual image is formed at infinity. The rays defining the image make an angle θ_{ey} with the telescope axis.

The angular magnification m_θ of the telescope is $\theta_{\text{ey}}/\theta_{\text{ob}}$. For paraxial rays (rays close to the axis) we can write $\theta_{\text{ob}} = h'/f_{\text{ob}}$ and $\theta_{\text{ey}} = h'/f_{\text{ey}}$, which gives

$$m_\theta = -\frac{f_{\text{ob}}}{f_{\text{ey}}}, \tag{40-34}$$

where the negative sign indicates an inverted final image.

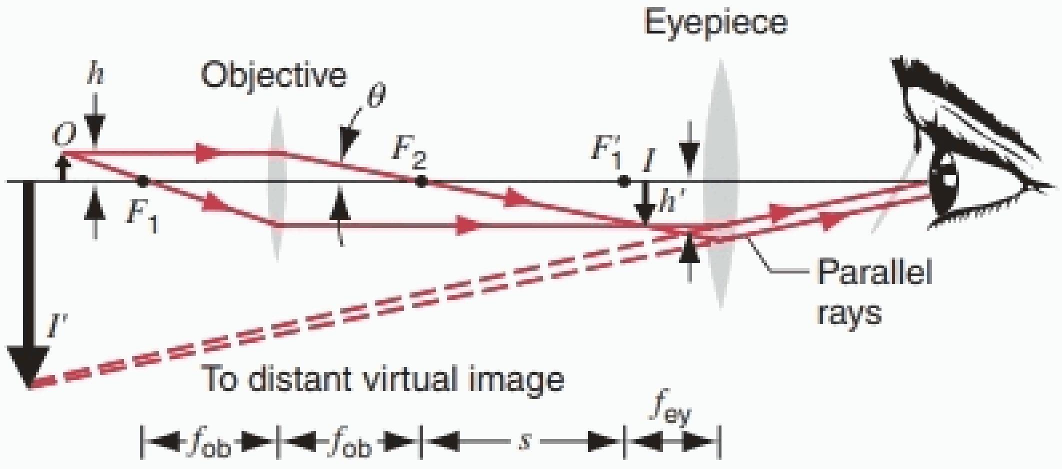


FIGURE 40-29. A thin-lens version of a microscope (not drawn to scale).

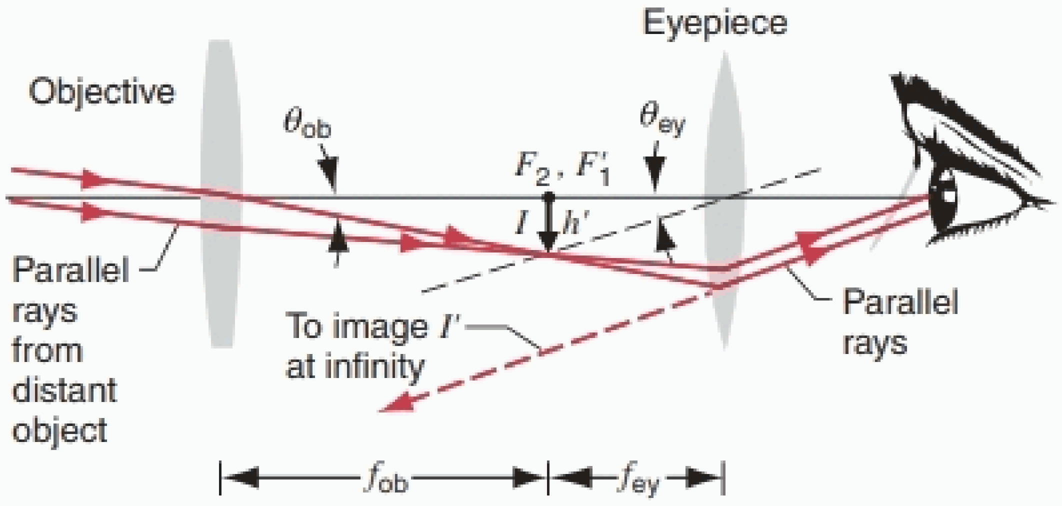


FIGURE 40-30. A thin-lens version of a refracting telescope (not drawn to scale).

Magnification is only one of the design factors of an astronomical telescope and is indeed easily achieved. A good telescope needs *light-gathering power,* which determines how bright the image is. This is important when viewing faint objects such as distant galaxies and is accomplished by making the objective lens diameter as large as possible. *Field of view* is another important parameter. An instrument designated for galactic observation (narrow field of view) must be quite different from one designed for the observation of meteors (wide field of view). The telescope designer must also take account of lens and mirror aberrations including *spherical aberration* (that is, lenses and mirrors with truly spherical surfaces do not form sharp images) and *chromatic aberration* (that is, for simple lenses the index of refraction and thus the focal length vary with wavelength so that fuzzy images are formed, displaying unnatural colors).

To build refracting telescopes of larger diameters (for better light-gathering efficiency), we must also make the lenses thicker, which increases the distortions and aberrations caused by the lens. The largest refracting telescopes, which were built around the end of the 19th century, have lenses about 1 m in diameter. Reflecting telescopes, in which the objective element is a mirror rather than a lens, do not suffer from these distortions, because the light reflects from the front surface of the mirror. The largest single reflecting telescopes have diameters around 10 m, and thus have about 100 times the light-gathering capability of the largest refracting telescopes. Even larger reflecting telescopes can be constructed by combining the light from many individual mirrors into a single image.

Earth-bound optical telescopes are limited in their ability to form sharp images by atmospheric distortion; the natural turbulence in the atmosphere distorts the (nearly) plane wavefronts that reach the Earth from distant objects. One cure for this problem has been obtained through the development of *adaptive optics*; by sensing the atmospheric distortion, the shape of a flexible mirror can be modified to compensate for the distortion and thus produce a sharp image. An alternative way to eliminate the effects of the atmosphere is to place the telescope *above* the atmosphere. Figure 40-31 shows the Hubble Space Telescope, a reflecting telescope that was launched into Earth orbit by a space shuttle in 1990.

FIGURE 40-31. The Hubble Space Telescope.

MULTIPLE CHOICE

40-1 Image Formation by Mirrors and Lenses

1. Is there such a thing as a virtual object?
   (A) No
   (B) Yes, but only as an unphysical mathematical construct
   (C) Yes, it is when rays of light originate from a point that is not there.
   (D) Yes, it is when converging rays of light strike an optical device before they can meet at a point.

40-2 Plane Mirrors

2. What happens to your image when you walk away from a plane mirror?
   (A) The image moves away from the mirror and gets smaller.
   (B) The image moves away from the mirror but stays the same size.
   (C) The image stays the same distance from the mirror but gets smaller.
   (D) The image stays the same distance from the mirror and stays the same size.

3. What is the minimum size mirror required for a 6-ft-tall person to be able to see a "full length" image?
   (A) 6 ft (B) 4.5 ft (C) 3 ft
   (D) It depends on the distance of the person from the mirror.

40-3 Spherical Mirrors

4. The passenger side-view mirror on an automobile often has the notation: "Objects in mirror are closer than they appear." Is the image *really* farther away than the object?
   (A) Yes, the image is smaller and farther away than the object.
   (B) No, the image is smaller and closer than the object.
   (C) No, the image is larger and closer than the object.
   (D) Yes, the image is larger and farther away than the object.

5. No matter how far you stand from a certain mirror, your image appears upright. What type of mirror is this?
(A) Concave (B) Convex (C) Plane
(D) Either (B) or (C)
(E) There is not enough information to answer this question.

6. While standing in front of a certain mirror, you notice that your image appears enlarged. What type of mirror is this?
(A) Concave (B) Convex (C) Plane
(D) Either (A) or (B)
(E) There is not enough information to answer this question.

7. A mirror produces a real image at i from a real object at $o > i$. What can you conclude about the focal length of the mirror?
(A) $f < 0$ (B) $0 < f < i$
(C) $i < f < o$ (D) $o < f$

40-4 Spherical Refracting Surfaces

8. Light shines through a spherical air bubble underwater. What type of optical device does the bubble act like?
(A) Converging (B) Diverging (C) Planar

9. A diver in a glass bubble helmet looks at a fish as in Fig. 40-32.
(*a*) Where is the image of the fish relative to its true location?
(A) Closer to the diver
(B) Farther from the diver
(C) The same distance from the diver
(D) The answer depends on the location of the fish.
(*b*) What is the size of the image compared with the size of the fish?
(A) Larger (B) Smaller (C) The same size
(D) The answer depends on the location of the fish.

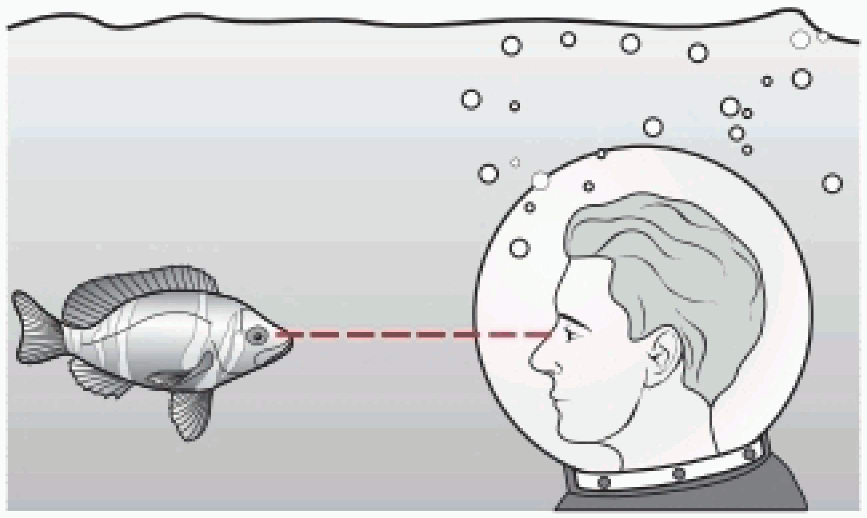


FIGURE 40-32. Multiple-choice question 9.

40-5 Thin Lenses

10. A nearsighted (see Exercise 40) person can clearly read a newspaper at a maximum distance of 1 m. What should be the focal length of this person's corrective lenses?
(A) More than 2 m (B) More than 1 m
(C) Less than 1 m (D) Less than 0.5 m

11. Figure 40-33 shows three people wearing glasses. Which person is nearsighted? (See Exercise 40.)

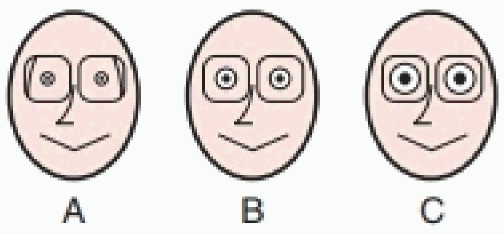


FIGURE 40-33. Multiple-choice question 11.

12. A lightbulb is a real object for a lens; an image of the bulb is projected onto a screen.
(*a*) What type of lens is this?
(A) Converging (B) Diverging
(C) Either (A) or (B)
(D) There is not enough information to answer this question.
(*b*) What type of image is produced?
(A) Real (B) Upright (C) Virtual
(D) Laterally inverted

13. A cartoon-character object is located to the left of a thin lens; a real image appears on the right. (See Fig. 40-34*a*.) Which picture in Fig. 40-34*b* best shows the image?

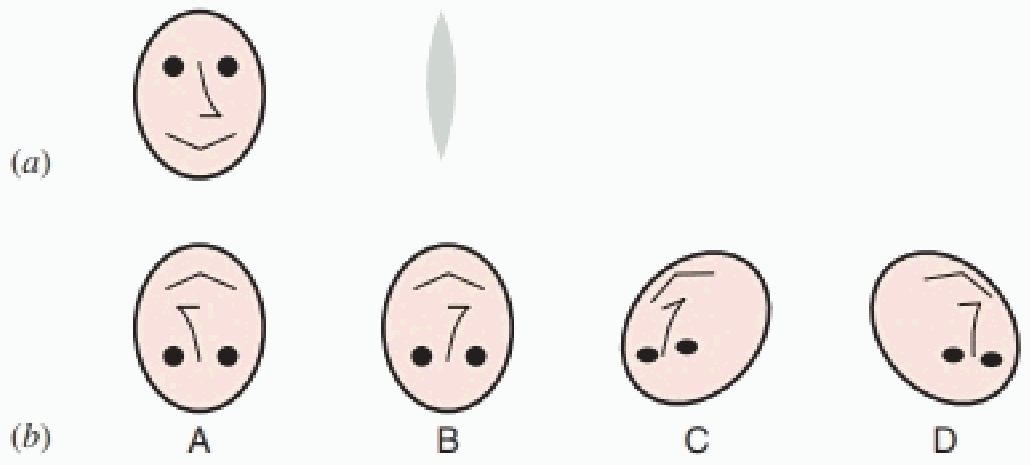


FIGURE 40-34. Multiple-choice question 13.

14. An opaque card is held over the lower half of a converging lens as shown in Fig. 40-35*a*. Which picture in Fig. 40-35*b* best shows the image that appears on the screen?

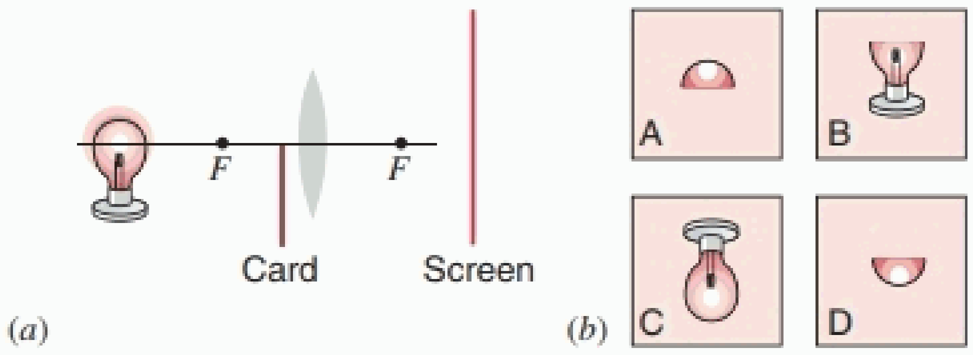


FIGURE 40-35. Multiple-choice question 14.

40-6 Optical Instruments

15. How far should a magnifying glass of focal length f be from an object to provide the largest *angular* magnification (as measured from the center of the lens) with an upright image?
(A) $o < f$ (B) $o \approx f$
(C) $f < o < 2f$ (D) $2f < o$
(E) It does not matter; the magnification is independent of o.

16. A magnifying glass of focal length f is held the optimal distance from an object to provide the largest magnification. At what distance should your eye be located from the magnifying glass to provide the largest angular magnification?
(A) Far from the glass
(B) Approximately f
(C) Very close to the glass
(D) It does not matter; the magnification is independent of your distance from the glass.

QUESTIONS

1. Can a virtual image be photographed by exposing a film at the location of the image? Explain.

2. At night, in a lighted room, you blow a smoke ring toward a window pane. If you focus your eyes on the ring as it approaches the pane, it will seem to go right through the glass into the darkness beyond. What is the explanation of this illusion?

3. In driving a car you sometimes see vehicles such as ambulances with letters printed on them in such a way that they read in the normal fashion when you look through the rear-view mirror. Print your name so that it may be so read.

4. We have seen that a single reflection in a plane mirror reverses right and left. When we drive down a highway, for example, the letters on the highway signs are reversed as seen through the rear-view mirror. And yet, as seen through this same mirror, you still seem to be driving down the right lane. Why does the mirror reverse the signs and not the lanes? Or does it? Discuss.

5. We all know that when we look into a mirror right and left are reversed. Our right hand will seem to be a left hand; if we part our hair on the left it will seem to be parted on the right, and so on. Can you think of a system of mirrors that would let us see ourselves as others see us? If so, draw it and prove your point by drawing some typical rays.

6. Device a system of plane mirrors that will let you see the back of your head. Trace the rays to prove your point.

7. What does "optical path length" mean? Can the optical path length ever be less than the geometrical path length? Ever be greater?

8. In many city buses a convex mirror is suspended over the door, in full view of the driver. Why not a plane or concave mirror?

9. Dentists and dental hygienists use a small mirror with a long handle attached to examine your teeth. Is the mirror concave, convex, or plane, and why?

10. Under what conditions will a spherical mirror, which may be concave or convex, form (*a*) a real image, (*b*) an inverted image, and (*c*) an image smaller than the object?

11. Can a virtual image be projected onto a screen?

12. You are looking at a dog through a glass window pane. Where is the image of the dog? Is it real or virtual? Is it upright or inverted? What is the magnification? (Hint: Think of the window pane as the limiting case of a thin lens in which the radii of curvature have been allowed to become infinitely large.)

13. In some cars the right (passenger) side-view mirror bears the notation: "Objects in the mirror are closer than they appear." What feature of the mirror requires this warning? What advantages does the mirror have to compensate for this disadvantage? Do cars viewed in this mirror appear to be moving faster or slower than they would if viewed in a plane mirror?

14. Are the objects in Question 13 really closer than they appear? Where is the image?

15. We have all seen TV pictures of a baseball game shot from a camera located somewhere behind second base. The pitcher and the batter are about 60 ft apart but they look much closer on the TV screen. Why are the images viewed through a telephoto lens foreshortened in this way?

16. An unsymmetrical thin lens forms an image of a point object on its axis. Is the image location changed if the lens is reversed?

17. Why has a lens two focal points and a mirror only one?

18. Under what conditions will a thin lens, which may be converging or diverging, form (*a*) a real image, (*b*) an inverted image, and (*c*) an image smaller than the object?

19. A diver wants to use an air-filled plastic bag as a converging lens for underwater visibility. Sketch a suitable cross section for the bag.

20. In connection with Fig. 40-23*a*, all rays originating on the same wavefront in the incident wave have the same optical path to the image point. Discuss this in connection with Fermat's principle (see Chapter 39).

21. Why does chromatic aberration occur in simple lenses but not in mirrors?

22. Consider various lens aberrations. Is it possible in principle to make a lens free of all aberrations (for example, by grinding the surfaces) when focusing monochromatic light?

23. A concave mirror and a converging lens have the same focal length in air. Do they have the same focal length when immersed in water? If not, which has the greater focal length?

24. Under what conditions will a thin lens have a lateral magnification (*a*) of -1 and (*b*) of $+1$?

25. How does the focal length of a thin glass lens for blue light compare with that for red light, assuming the lens is (*a*) diverging and (*b*) converging?

26. Does the focal length of a lens depend on the medium in which the lens is immersed? Is it possible for a given lens to act as a converging lens in one medium and a diverging lens in another medium?

27. Are the following statements true for a glass lens in air? (*a*) A lens that is thicker at the center than at the edges is a converging lens for parallel light. (*b*) A lens that is thicker at the edges than at the center is a diverging lens for parallel light.

28. Under what conditions would the lateral magnification ($m = -i/o$) for lenses and mirrors become infinite? Is there any practical significance to such a condition?

29. Why is the magnification of a simple magnifier (see the derivation leading to Eq. 40-31) defined in terms of angles rather than image/object size?

30. Ordinary spectacles do not magnify but a simple magnifier does. What then is the function of spectacles?

31. The *f-number* of a camera lens (see Problem 13) is its focal length divided by its aperture (effective diameter). Why is this useful to know in photography? How can the *f*-number of the lens be changed? How is exposure time related to *f*-number?

32. A magnifying glass of small focal length allows finer detail to be examined than does one of long focal length. Explain.

33. Estimate the greatest distance at which the human eye can read the headlines of a newspaper.

34. Does it matter whether (*a*) an astronomical telescope; (*b*) a compound microscope; (*c*) a simple magnifier; (*d*) a camera, including a TV camera; or (*e*) a projector, including a slide projector, produces upright or inverted images? What about real or virtual images?

35. The unaided human eye produces a real but inverted image on the retina. (*a*) Why then do we not perceive objects such as people and trees as upside down? (*b*) We do not, of course, but suppose that we wore special glasses so that we did. If you then turned this book upside down, could you read this question with the same facility that you do now?

36. Which of the following—a converging lens, a diverging lens, a concave mirror, a convex mirror, or a plane mirror—is used: (*a*) As a magnifying glass? (*b*) As the reflector in the lamphouse of a slide projector? (*c*) As the objective of a reflecting telescope? (*d*) In a kaleidoscope? (*e*) As the eyepiece of opera glasses? (*f*) To obtain a wider rear view from the driver's seat in a car?

37. What properties of a lens would make it a good burning glass (a lens that, aimed at the Sun, will quickly ignite paper or twigs placed behind it)?

38. In William Golding's *Lord of the Flies,* the character Piggy uses his glasses to focus the Sun's rays and kindle a fire. Later, the boys abuse Piggy and break his glasses. He is unable to identify them at close range because he is nearsighted. Find the flaw in this narrative. (Boston Globe, December 17, 1985, Letters.)

39. Explain the function of the objective lens of a microscope. Why use an objective lens at all? Why not just use a very powerful magnifier?

40. Why do astronomers use optical telescopes in looking at the sky? After all, the stars are so far away that they still appear to be points of light, without any detail discernible.

41. A watchmaker uses diverging eyeglasses for driving, no glasses for reading, and converging glasses in occupational work. Is the watchmaker nearsighted or farsighted? Explain. (See Exercise 40.)

42. Why are all recent large astronomical telescopes of the reflecting rather than the refracting variety? Think of mechanical mounting problems for lenses and mirrors, the difficulty of shaping (that is, "figuring") the various optical surfaces involved, problems with small flaws in the optical glass blanks used to make lenses and mirrors, and so on.

43. Explain why (*a*) ultraviolet light is sometimes used to illuminate objects under a microscope, (*b*) blue filters are sometimes used to photograph a star seen through a telescope, and (*c*) infrared light is often used to get greater clarity in landscape photographs.

EXERCISES

40-1 Image Formation by Mirrors and Lenses

40-2 Plane Mirrors

1. You are standing in front of a large plane mirror, contemplating your image. If you move toward the mirror at speed v, at what speed does your image move toward you? Report this speed both (*a*) in your own reference frame and (*b*) in the reference frame of the room in which the mirror is at rest.

2. A small object is 10 cm in front of a plane mirror. If you stand behind the object, 30 cm from the mirror, and look at its image, for what distance must you focus your eyes?

3. Prove that if a plane mirror is rotated through an angle α, the reflected beam is rotated through an angle 2α. Show that this result is reasonable for $\alpha = 45°$.

4. Figure 40-36 shows (top view) that Bernie B is walking directly toward the center of a vertical mirror M. How close to the mirror will he be when Sarah S is just able to see him? Take $d = 3.0$ m.

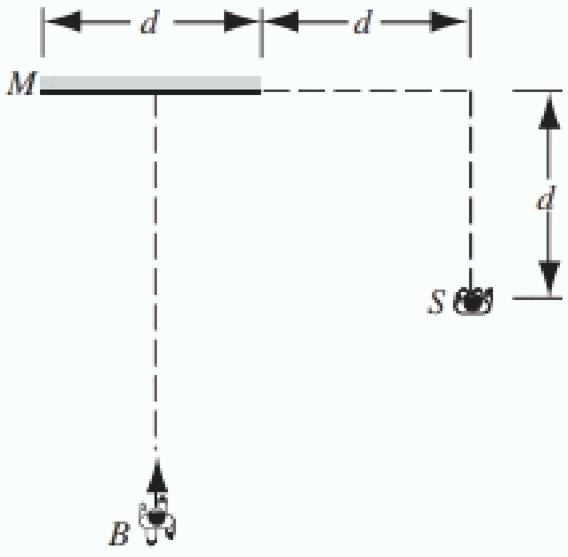


FIGURE 40-36. Exercise 4.

5. A small object O is placed one-third of the way between two parallel plane mirrors as in Fig. 40-37. Trace appropriate bundles of rays for viewing the four images that lie closest to the object.

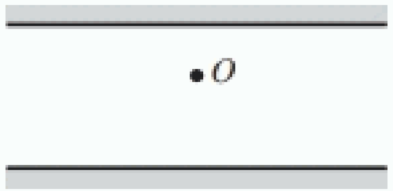


FIGURE 40-37. Exercise 5.

6. In Fig. 40-4, you rotate the mirror 30° counterclockwise about its bottom edge, leaving the point object O in place. Is the image point displaced? If so, where is it? Can the eye still see the image without being moved? Sketch a figure showing the new situation.

7. Figure 40-38 shows a small lightbulb suspended 250 cm above the surface of the water in a swimming pool. The water is 186 cm deep and the bottom of the pool is a large mirror. Where is the image of the lightbulb when viewed from near normal incidence?

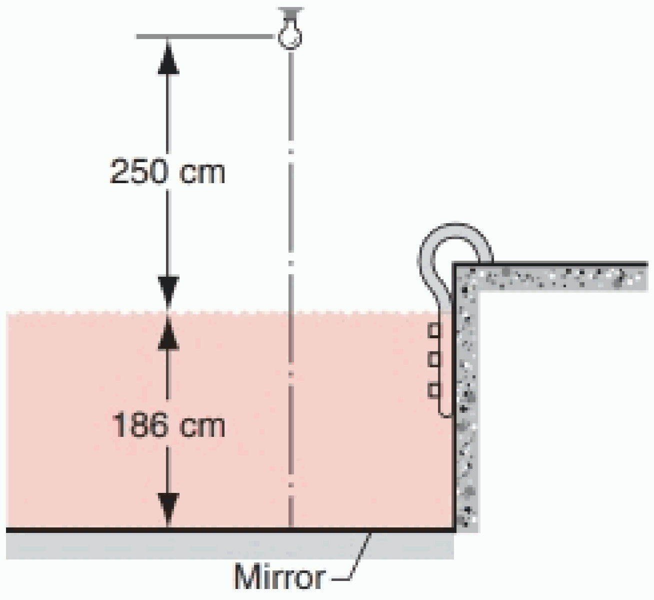


FIGURE 40-38. Exercise 7.

8. Two plane mirrors make an angle of 90° with each other. What is the largest number of images of an object placed between them that can be seen by a properly placed eye? The object does not lie on the mirror bisector.

9. A point object is 10 cm away from a plane mirror while the eye of an observer (pupil diameter 5.0 mm) is 24 cm away. Assuming both the eye and the point to be on the same line perpendicular to the surface, find the area of the mirror used in observing the reflection of the point.

10. Solve Exercise 8 if the angle between the mirrors is (*a*) 45°, (*b*) 60°, and (*c*) 120°, the object always being placed on the bisector of the mirrors.

11. How many images of yourself can you see in a room whose ceiling and two adjacent walls are mirrors? Explain.

40-3 Spherical Mirrors

12. Redraw Fig. 40-39 on a large sheet of paper and trace carefully the reflected rays, using the law of reflection. Is a point focus formed? Discuss.

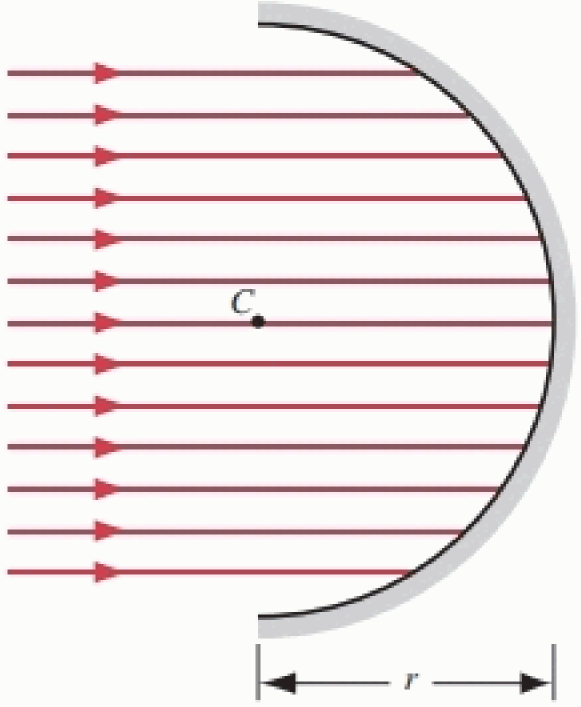


FIGURE 40-39. Exercise 12.

13. A concave shaving mirror has a radius of curvature of 35 cm. It is positioned so that the image of a man's face is 2.7 times the size of his face. How far is the mirror from the man's face?

14. Complete Fig. 40-40 to show the path followed by each of the rays of light from the point object O after they reflect from the mirror. The focal length is 10.0 cm, and the object distance is 15.0 cm.

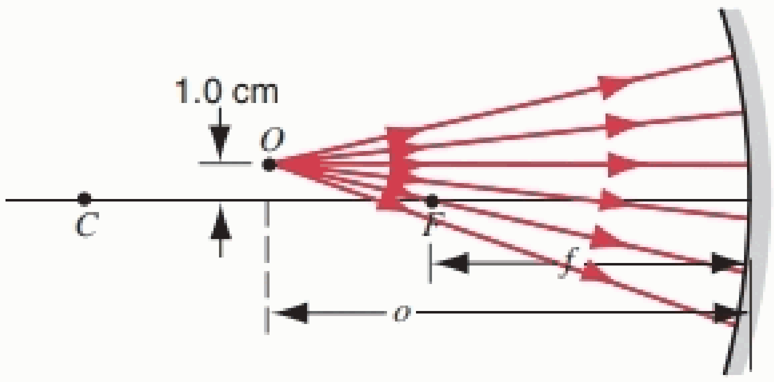


FIGURE 40-40. Exercise 14.

15. Fill in Table 40-1, where each column refers to a spherical or plane mirror and a real object. Check your results by ray-tracing. Distances are in centimeters; if a number has no positive or negative sign in front of it, find the correct sign.

16. Repeat the derivation leading to Eq. 40-4 using the geometry of Fig. 40-15 for the convex mirror, and show that Eq. 40-4 is valid in this case only if i and r are taken to be negative.

40-4 Spherical Refracting Surfaces

17. Figure 40-41 shows the cross section of a hollow glass tube of internal radius r, external radius R, and index of refraction n. (*a*) Convince yourself that the ray ABC shown defines the apparent internal radius r^* as seen from the side. (*b*) Show that $r^* = nr$, independent of R.

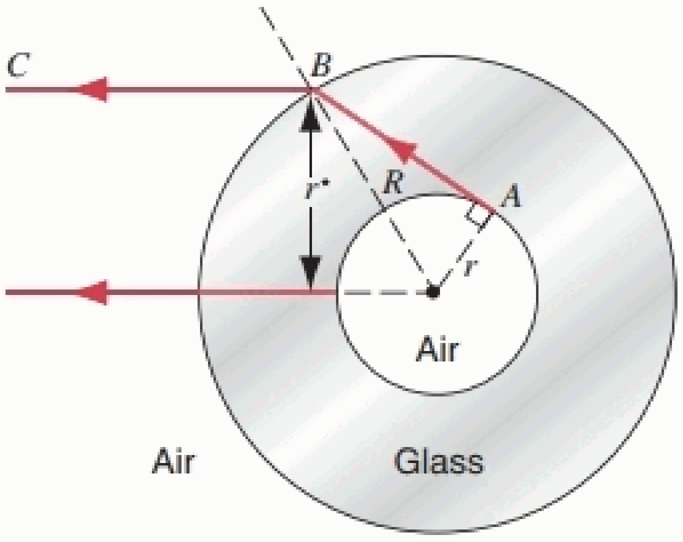


FIGURE 40-41. Exercise 17.

TABLE 40-1 Exercise 15.

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
Type	Concave						Convex	
f (cm)	20		+ 20			20		
r (cm)					− 40		40	
i (cm)					− 10		4.0	
o (cm)	+ 10	+ 10	+ 30	+ 60				+ 24
m		+ 1.0		− 0.50		+ 0.10		0.50
Real image?		No						
Upright image?								No

TABLE 40-2 Exercise 18.

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
n_1	1.0	1.0	1.0	1.0	1.5	1.5	1.5	1.5
n_2	1.5	1.5	1.5		1.0	1.0	1.0	
o (cm)	+ 10	+ 10		+ 20	+ 10		+ 70	+ 100
i (cm)		− 13	+ 600	− 20	− 6.0	− 7.5		+ 600
r (cm)	+ 30		+ 30	− 20		− 30	+ 30	− 30
Real image?								

18. Fill in Table 40-2, where each column refers to a spherical surface separating two media with different indices of refraction. Distances are measured in centimeters. The object is real in all cases. Draw a figure for each situation and construct the appropriate rays graphically. Assume a point object.

19. A parallel beam of light from a laser falls on a solid, transparent sphere of index of refraction *n*, as shown in Fig. 40-42. (*a*) Show that the beam cannot be brought to a focus at the back of the sphere unless the beam width is small compared with the radius of the sphere. (*b*) If the condition in (*a*) is satisfied, what is the index of refraction of the sphere? (*c*) What index of retraction, if any, will focus the beam at the center of the sphere?

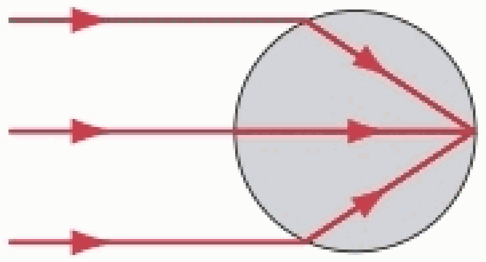


FIGURE 40-42. Exercise 19.

40-5 Thin Lenses

20. Complete Fig. 40-43 to show the path followed by each of the rays of light from the point object *O* after they pass through the lens. The focal length is 10.0 cm, and the object distance is 6.0 cm.

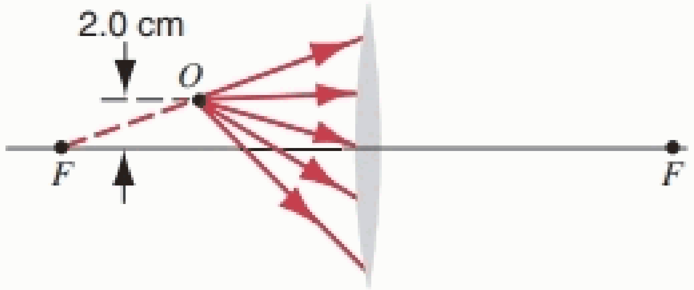


FIGURE 40-43. Exercise 20.

21. An object is 20 cm to the left of a thin diverging lens having a focal length of − 30 cm. Where is the image formed? Obtain the image position both by calculation and also from a ray diagram.

22. A double-convex lens is to be made of glass with an index of refraction of 1.5. One surface is to have twice the radius of curvature of the other and the focal length is to be 60 mm. Find the radii.

23. Suppose that you focus an image of the Sun on a screen, using a thin lens whose focal length is 27 cm. Find the diameter of the image. (See Appendix C for needed data on the Sun.)

24. A lens is made of glass having an index of refraction of 1.5. One side of the lens is flat and the other convex wih a radius of curvature of 20 cm. (*a*) Find the focal length of the lens. (*b*) If an object is placed 40 cm to the left of the lens, where will the image be located?

25. You have a supply of flat glass disks ($n = 1.5$) and a lens-grinding machine that can be set to grind radii of curvature of either 40 cm or 60 cm. You are asked to prepare a set of six lenses like those shown in Fig. 40-44. What will be the focal length of each lens? (Note: Where you have a choice of radii of curvature, select the smaller one.)

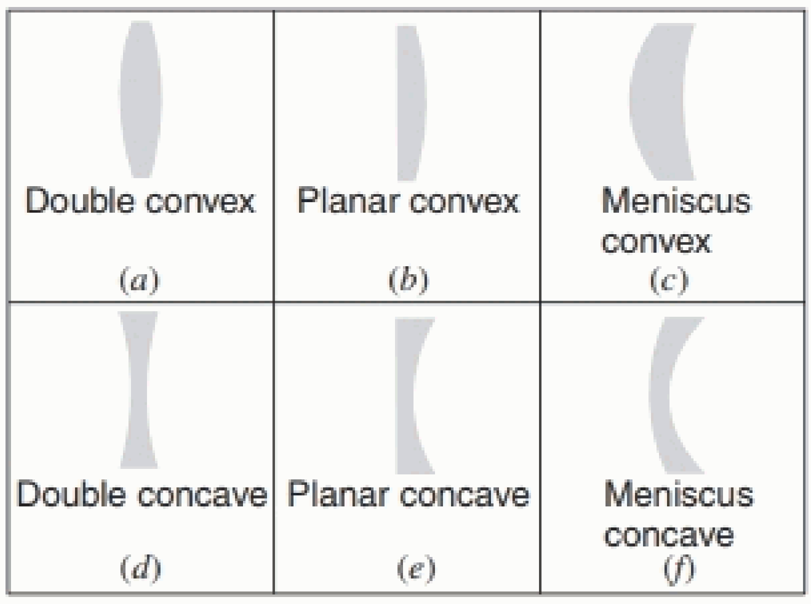


FIGURE 40-44. Exercise 25.

26. An object is placed at the center of curvature of a double concave lens, both of whose radii of curvature have the same magnitude. (*a*) Find the image distance in terms of the radius of curvature *r* and the index of refraction *n* of the glass. (*b*) Describe the nature of the image. (*c*) Verify your result with a ray diagram.

27. Equation 40-15 is called the *Gaussian* form of the thin-lens formula. Another form of this formula, the *Newtonian* form, is obtained by considering the distance *x* from the object to the first focal point and the distance *x*′ from the second focal point to the image. Show that

$$xx' = f^2.$$

TABLE 40-3 Exercise 28.

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
Type	Converging								
f (cm)	10	+ 10	10	10					
r_1 (cm)					+ 30	− 30	− 30		
r_2 (cm)					− 30	+ 30	− 60		
i (cm)									
o (cm)	+ 20	+ 5.0	+ 5.0	+ 5.0	+ 10	+ 10	+ 10	+ 10	+ 10
n					1.5	1.5	1.5		
m			>1	<1				0.50	0.50
Real image?									Yes
Upright image?								Yes	

28. To the extent possible, fill in Table 40-3, where each column refers to a thin lens. If a quantity cannot be calculated, write "X." Distances are in centimeters; if a number (except in row n) has no positive or negative sign in front of it, find the correct sign. Draw a figure for each situation and construct the appropriate rays graphically. The object is real in all cases.

29. An illuminated slide is mounted 44 cm from a screen. How far from the slide must a lens of focal length 11 cm be placed in order to focus an image on the screen?

30. An illuminated arrow forms a real inverted image of itself at a distance $d = 40.0$ cm, measured along the optic axis of a lens; see Fig. 40-45. The image is only half the size of the object. (*a*) What kind of lens must be used to produce this image? (*b*) How far from the object must the lens be placed? (*c*) What is the focal length of the lens?

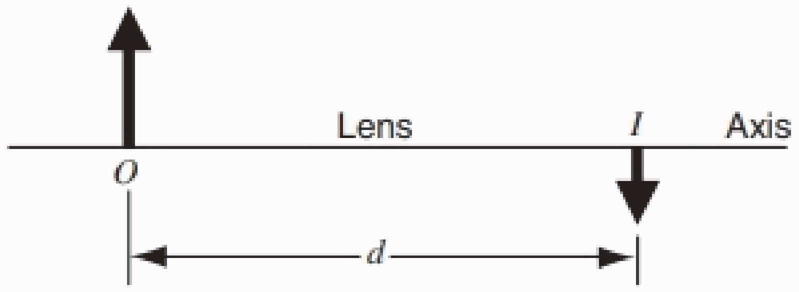


FIGURE 40-45. Exercise 30.

31. A converging lens with a focal length of $+20$ cm is located 10 cm to the left of a diverging lens having a focal length of -15 cm. If a real object is located 40 cm to the left of the first lens, locate and describe completely the image formed.

32. (*a*) Show that a thin converging lens of focal length f followed by a thin diverging lens of focal length $-f$ will bring parallel light to a focus beyond the second lens provided that the separation of the lenses L satisfies $0 < L < f$. (*b*) Does this property change if the lenses are interchanged? (*c*) What happens when $L = 0$?

33. An object is placed 1.12 m in front of a converging lens of focal length 58.0 cm, which is 1.97 m in front of a plane mirror. (*a*) Where is the final image, measured from the lens, that would be seen by an eye looking toward the mirror through the lens? (*b*) Is the final image real or virtual? (*c*) Is the final image upright or inverted? (*d*) What is the lateral magnification?

34. An object is 20.0 cm to the left of a lens with a focal length of $+10.0$ cm. A second lens of focal length $+12.5$ cm is 30.0 cm to the right of the first lens. (*a*) Using the image formed by the first lens as the object for the second, find the location and relative size of the final image. (*b*) Verify your conclusions by drawing the lens system to scale and constructing a ray diagram. (*c*) Describe the final image.

40-6 Optical Instruments

35. The angular magnification of an astronomical telescope in normal adjustment is 36, and the diameter of the objective lens is 72 mm. What is the minimum diameter of the eyepiece required to collect all the light entering the objective lens from a distant point source on the axis of the instrument?

36. Some of the first "microscopes" were really single-lens magnifying glasses made by Leeuwenhoek. These devices had magnifications as high as 200. Assume that these were made from a glass ($n = 1.5$) lens with two surfaces of equal curvature R. (*a*) Find R. (*b*) Estimate the maximum diameter of these lenses.

37. Figure 40-46*a* suggests a normal human eye. Parallel rays entering a relaxed eye gazing at infinity produce a real, inverted image on the retina. The eye thus acts as a converging

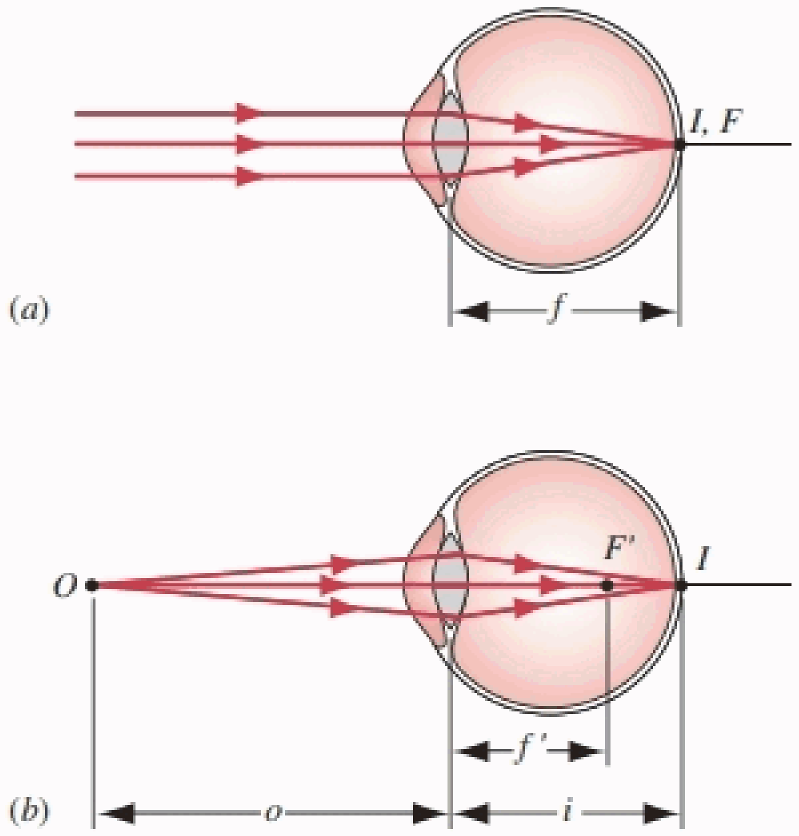


FIGURE 40-46. Exercise 37.

lens. Most of the refraction occurs at the outer surface of the eye, the *cornea.* Assume a focal length f for the eye of 2.50 cm. In Fig. 40-46*b* the object is moved in to a distance $o = 36.0$ cm from the eye. To form an image on the retina, the effective focal length of the eye must be reduced to f'. This is done by the action of the ciliary muscles that change the shape of the lens and thus the effective focal length of the eye. (*a*) Find f' from the above data. (*b*) Would the effective radii of curvature of the lens become larger or smaller in the transition from Fig. 40-46*a* to 40-46*b*? (In the figure the structure of the eye is only roughly suggested and Fig. 40-46*b* is not to scale.)

38. A microscope of the type shown in Fig. 40-29 has a focal length for the objective lens of 4.2 cm and for the eyepiece lens of 7.7 cm. The distance between the lenses is 25 cm. (*a*) What is the distance s in Fig. 40-29? (*b*) To reproduce the conditions of Fig. 40-29, how far beyond F_1 in that figure should the object be placed? (*c*) What is the lateral magnification m of the objective lens? (*d*) What is the angular magnification m_θ of the eyepiece? (*e*) What is the overall magnification M of the microscope?

39. In a compound microscope, the object is 12.0 mm from the objective lens. The lenses are 285 mm apart and the intermediate image is 48.0 mm from the eyepiece. What magnification is produced?

40. In an eye that is *farsighted,* the eye focuses parallel rays so that the image would form behind the retina, as in Fig. 40-47*a*. In an eye that is *nearsighted,* the image is formed in front of the retina, as in Fig. 40-47*b*. (*a*) How would you design a corrective lens for each eye defect? Make a ray diagram for each case. (*b*) If you need spectacles only for reading, are you nearsighted or farsighted? (*c*) What is the function of bifocal spectacles, in which the upper parts and lower parts have different focal lengths?

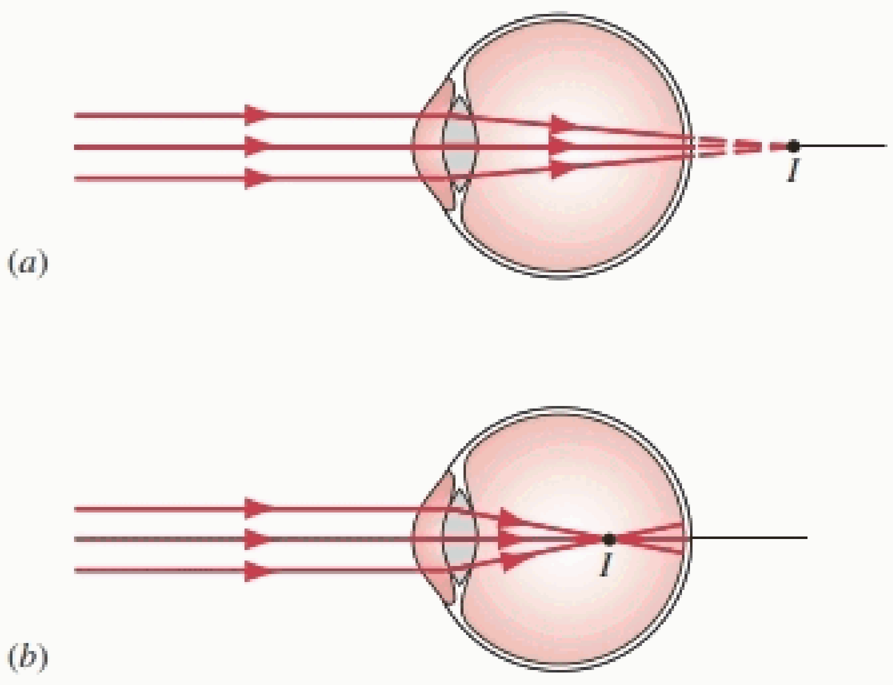


FIGURE 40-47. Exercise 40.

PROBLEMS

1. You put a point source of light S a distance d in front of a screen A. How is the intensity at the center of the screen changed if you put a mirror M a distance d behind the source, as in Fig. 40-48? (Hint: Recall the variation of intensity with distance from a point source of light.)

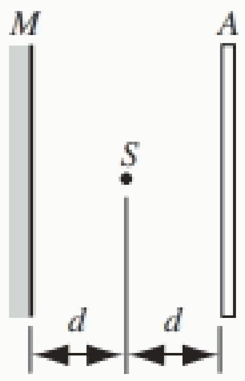


FIGURE 40-48. Problem 1.

2. (*a*) A luminous point is moving at speed v_o toward a spherical mirror, along its axis. Show that the speed at which the image of this point object is moving is given by

$$v_i = -\left(\frac{r}{2o - r}\right)^2 v_o.$$

(*b*) Assume that the mirror is concave, with $r = 15$ cm, and that $v_o = 5.0$ cm/s. Find the speed of the image if the object is far outside the focal point ($o = 75$ cm). (*c*) If it is close to the focal point ($o = 7.7$ cm). (*d*) If it is very close to the mirror ($o = 0.15$ cm).

3. A short linear object of length L lies on the axis of a spherical mirror, a distance o from the mirror. (*a*) Show that its image will have a length L', where

$$L' = L\left(\frac{f}{o - f}\right)^2.$$

(*b*) Show that the *longitudinal magnification* m' $(= L'/L)$ is equal to m^2, where m is the lateral magnification.

4. A narrow, parallel incident beam of light falls, from the left, on a solid glass sphere at normal incidence. The radius of the sphere is R and its index of refraction is $n < 2$. Find the distance of the image from the right edge of the sphere.

5. Show that the focal length f for a thin lens whose index of refraction is n and that is immersed in a fluid whose index of refraction is n' is given by

$$\frac{1}{f} = \frac{n - n'}{n'}\left(\frac{1}{r_1} - \frac{1}{r_2}\right).$$

6. Let $y = o/|f|$ and $y' = i/|f|$. Show that Eqs. 40-4 and 40-15 can be written as

$$\frac{1}{y} + \frac{1}{y'} = \pm 1.$$

Plot a graph of y vs. y' when the above sum is $+1$, and another graph for -1. Use -4 to $+4$ as the range for both y and y'. (*a*) Label the graphs that correspond to concave mirrors, convex mirrors, converging lenses, and diverging lenses. (*b*) Identify the portions of each graph that correspond to real objects, virtual objects, real images, and virtual images. (*c*) Mark the points on each curve when the magnification is ± 2, ± 1, and ± 0.5.

7. A lumious object and a screen are a fixed distance D apart. (a) Show that a converging lens of focal length f will form a real image on the screen for two positions that are separated by

$$d = \sqrt{D(D - 4f)}.$$

(b) Show that the ratio of the two image sizes for these two positions is

$$\left(\frac{D - d}{D + d}\right)^2.$$

8. Show that the distance between a real object and its real image formed by a thin converging lens is always greater than or equal to four times the focal length of the lens.

9. Two converging lenses, with focal lengths f_1 and f_2, are positioned a distance $f_1 + f_2$ apart, as shown in Fig. 40-49. Arrangements like this are called *beam expanders* and are often used to increase the diameters of light beams from lasers. (a) If W_1 is the incident beam width, show that the width of the emerging beam is $W_2 = (f_2/f_1)W_1$. (b) Show how a combination of one diverging and one converging lens can also be arranged as a beam expander. Incident rays parallel to the axis should exit parallel to the axis. (c) Calculate the ratio of the intensity of the beam emerging from the beam expander to the intensity of the laser beam.

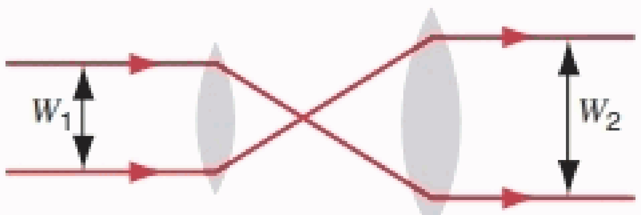


FIGURE 40-49. Problem 9.

10. A thin, flat plate of partially reflecting glass is a distance b from a convex mirror. A point source of light S is placed a distance a in front of the plate (see Fig. 40-50) so that its image in the partially reflecting plate coincides with its image in the mirror. If $b = 7.50$ cm and the focal length of the mirror is $f = -28.2$ cm, find a and draw the ray diagram.

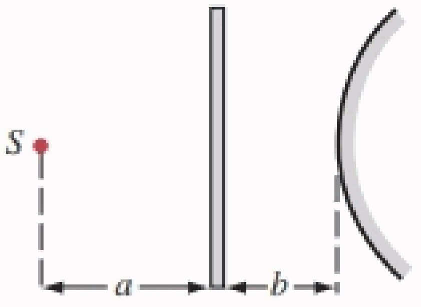


FIGURE 40-50. Problem 10.

11. Two thin lenses of focal lengths f_1 and f_2 are in contact. Show that they are equivalent to a single thin lens with a focal length given by

$$f = \frac{f_1 f_2}{f_1 + f_2}.$$

12. An upright object is placed a distance in front of a converging lens equal to twice the focal length f_1 of the lens. On the other side of the lens is a converging mirror of focal length f_2, separated from the lens by a distance $2(f_1 + f_2)$; see Fig. 40-51. (a) Find the location, nature, and relative size of the final image, as seen by an eye looking toward the mirror through the lens. (b) Draw the appropriate ray diagram.

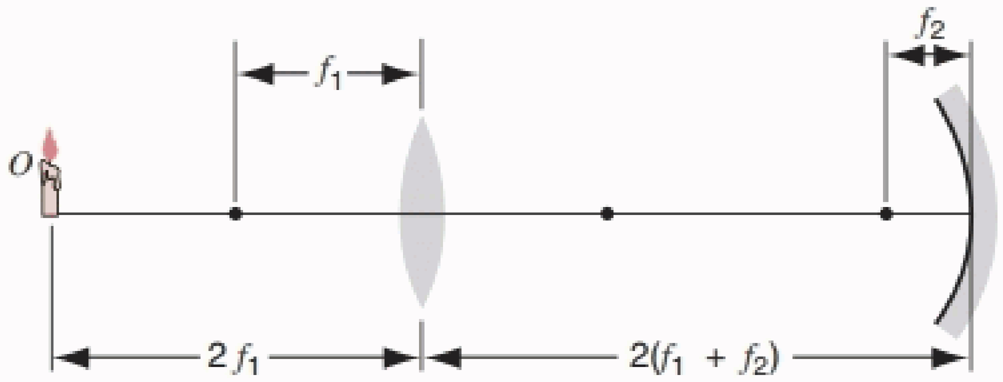


FIGURE 40-51. Problem 12.

13. Figure 40-52 shows an idealized camera focused on an object at infinity. A real, inverted image I is formed on the film, with the image distance i equal to the (fixed) focal length f (=5.0 cm, say) of the lens system. In Fig. 40-52b the object O is closer to the camera, the object distance o being, say, 100 cm. To focus an image I on the film, we must extend the lens away from the camera (why?). (a) Find i' in Fig. 40-52b. (b) By how much must the lens be moved? Note that the camera differs from the eye (see Exercise 37) in this respect. In the camera, f remains constant and the image distance i must be adjusted by moving the lens. For the eye, the image distance i remains constant and the focal length f is adjusted by distorting the lens. Compare Fig. 40-46 and Fig. 40-52 carefully. (c) The focal length of a small camera is 50 mm and the focusing range extends from 1.2 m out to infinity. Find the range of movement necessary between lens and film.

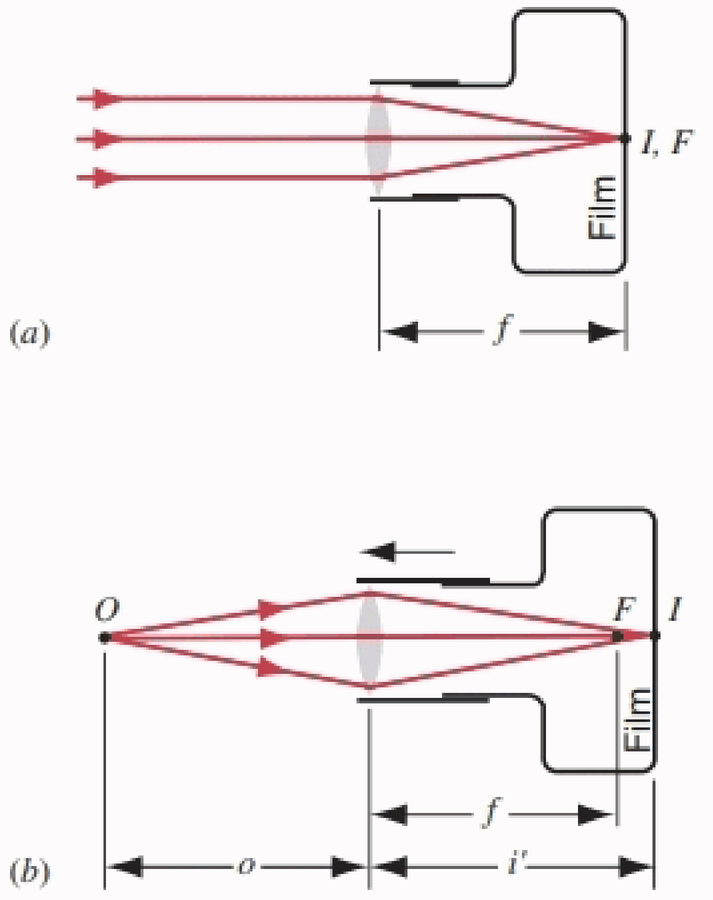


FIGURE 40-52. Problem 13.

14. The power P of a lens is defined by $P = 1/f$, where f is the focal length. The unit of power is the diopter, where 1 diopter = 1/meter. (a) Why is this a reasonable definition to use for lens power? (b) Show that the net power of two lenses in contact is given by $P = P_1 + P_2$, where P_1 and P_2 are the powers of the separate lenses. (Hint: See Problem 11.)

15. A photographer stands 44.5 m from a railroad track, the line of vision being perpendicular to the tracks. A train passes at 135 km/h and the photographer takes a picture. Using a camera with focal length 3.6 cm, find the maximum exposure time so that the blurring of the image on the film does not exceed 0.75 mm.

16. Isaac Newton, having convinced himself (erroneously as it turned out) that chromatic aberration was an inherent property of refracting telescopes, invented the reflecting telescope, shown schematically in Fig. 40-53. He presented his second model of this telescope, which has a magnifying power of 38, to the Royal Society, which still has it. In Fig. 40-53, incident light falls, closely parallel to the telescope axis, on the objective mirror *M*. After reflection from small mirror *M′* (the figure is not to scale), the rays form a real, inverted image in the focal plane through *F*. This image is then viewed through an eyepiece. (*a*) Show that the angular magnification m_θ is also given by Eq. 40-34, or

$$m_\theta = -f_{\text{ob}}/f_{\text{ey}},$$

where f_{ob} is the focal length of the objective mirror and f_{ey} is that of the eyepiece. (*b*) The 200-in. mirror in the reflecting telescope at Mt. Palomar Observatory in California has a focal length of 16.8 m. Estimate the size of the image formed in the focal plane of this mirror when the object is a meter stick 2.0 km away. Assume parallel incident rays. (*c*) The mirror of a different reflecting astronomical telescope has an effective radius of curvature ("effective" because such mirrors are ground to a parabolic rather than a spherical shape, to eliminate spherical aberration defects) of 10 m. To give an angular magnification of 200, what must be the focal length of the eyepiece?

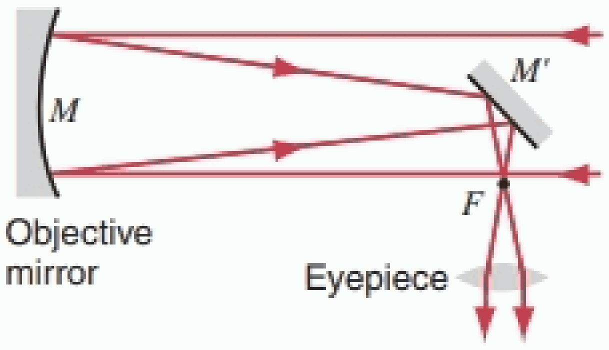


FIGURE 40-53. Problem 16.

COMPUTER PROBLEM

1. For a real concave spherical mirror, the light from parallel rays is not all reflected to the focal point. Assume that you have a concave spherical mirror with a 1.0-m radius of curvature. (*a*) For incident rays parallel to the axis, write a computer program that gives the point where the reflected ray crosses the axis as a function of the distance *d* of the incident ray from the axis. (*b*) Use your program to generate a series of reflected rays for incident distances of $d = 0, 5, 10, 15, \ldots$ cm, up to $d = 1.0$ m; draw these rays on a single diagram. The outer envelope of these rays is called a *caustic*. You can see this pattern when bright light reflects from the inside of a spoon onto a sheet of paper. (See Exercise 12 and Fig. 40-39.)

CHAPTER 41

INTERFERENCE

In this chapter and the next one, we discuss interference and diffraction of light waves. These occur in the realm of physical optics (or wave optics), in which a light beam encounters obstacles or openings whose size is comparable to the wavelength of the light. In contrast to the behavior of lenses and mirrors, which are described by the methods of geometrical optics where we represented light as rays, the phenomena of physical optics depend on the wave nature of light. In fact, it is from these experiments that convincing proof of the wave nature of light was first obtained.

41-1 TWO-SOURCE INTERFERENCE

When otherwise identical waves from two sources overlap at a point in space, the combined wave intensity at that point can be greater or less than the intensity of either of the two waves. We call this effect *interference.* The interference can be either *constructive,* when the net intensity is greater than the individual intensities, or *destructive,* when the net intensity is less than the individual intensities. Whether the interference is constructive or destructive depends on the relative phase of the two waves.

Although any number of waves can in principle interfere, we consider here the interference of only two waves. We assume that the sources of the waves each emit at only a single wavelength. (The case of sources that emit waves of several wavelengths can be handled by considering the separate interferences of the individual component wavelengths.)

Figure 41-1 represents the time dependence of two identical waves that arrive at the same point P in space. In Fig. 41-1*a*, the two waves arrive *in phase;* that is, they line up crest to crest and valley to valley. The net effect at point P is due to the combination of the two waves, and the resultant wave would have twice the amplitude of the component waves. This is the condition of maximal constructive interference. The pattern of Fig. 41-1*a* would look the same if we shifted one of the waves by a full cycle (2π radians) or by any whole number of full cycles. We can therefore say that *maximal constructive interference of two waves occurs when their phase difference (in radians) is 0, 2π, 4π, ...*

In Fig. 41-1*b*, the waves line up crest to valley and valley to crest. This is the condition for complete destructive interference, in which the two waves cancel one another's effect at point P. To go from a situation of constructive interference

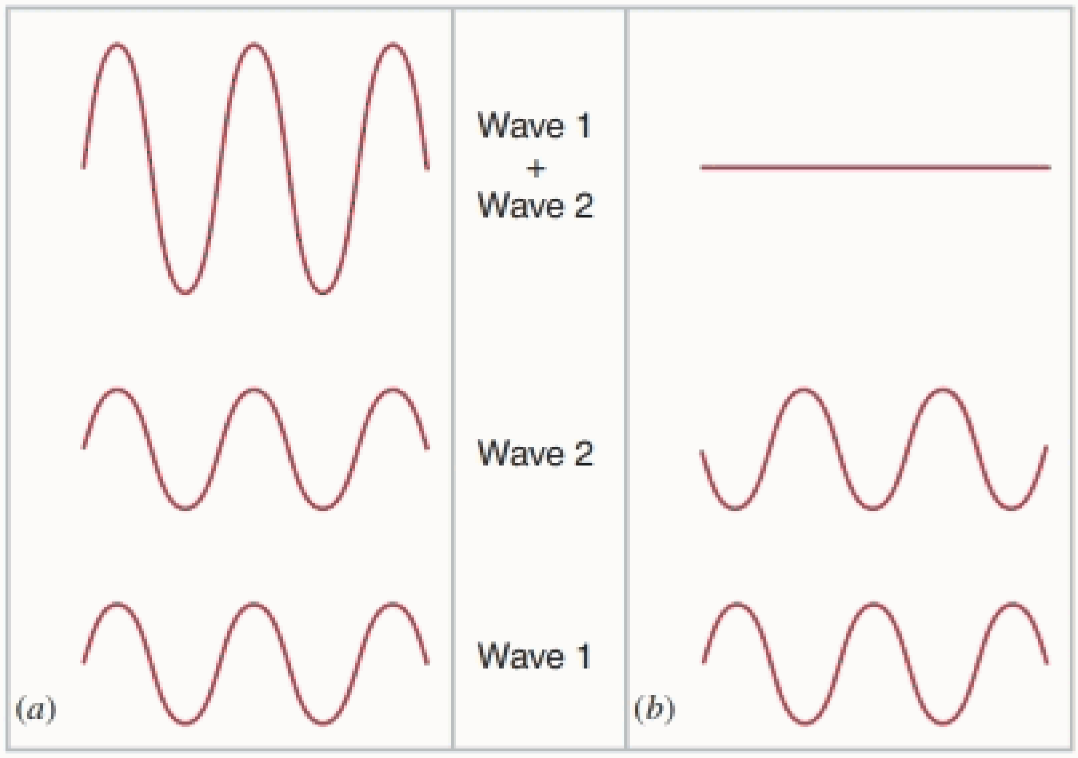


FIGURE 41-1. (*a*) Constructive interference of two waves that are in phase. (*b*) Destructive interference of two waves that are 180° out of phase.

(as in Fig. 41-1*a*) to one of destructive interference (Fig. 41-1*b*), we must shift one of the waves by one-half cycle (π radians), or by any whole number of full cycles plus a half cycle. That is, *complete destructive interference of two waves occurs when their phase difference (in radians) is* π, 3π, 5π, . . . ; equivalently, we say that the waves are *180° out of phase.*

We assume, for the time being, that the phase relationship between the two waves does not change with time. Such waves are said to be *coherent.* When coherent waves interfere, the intensity of the combined wave at any point in space does not change with time. Coherence, which is a necessary condition for interference to occur, is discussed in Section 41-3.

Two different light sources cannot in general be made coherent, because the emission of light by the atoms of one source is independent of the emission of the other source. The peaks and valleys of the waves from the two sources do not maintain a definite phase relationship, and the waves are said to be *incoherent.* To perform interference experiments with light, it is usually necessary to divide the light from a single source into two components and to treat each component as if it were emitted from an independent source of light. If we do this properly, the two components can be made to interfere. In this chapter we consider several schemes to create this division of the light wave.

If two sources emit waves throughout a region of space, then it is possible simultaneously to have constructive interference at some points and destructive interference at other points. That is, there may be some points in space where the distances from the observation point to the two sources is just right for the two waves to arrive in phase, as in Fig. 41-1*a*. These points are the locations of the *interference maxima.* There might be other locations where the distances to the two sources cause the two waves to arrive out of phase at the observation point, as in Fig. 41-1*b*. These are the *interference minima.* Figure 41-2 shows a pattern of maxima and minima produced by interfering water waves in a ripple tank. As we see in the next section, interfering light waves produce a similar pattern.

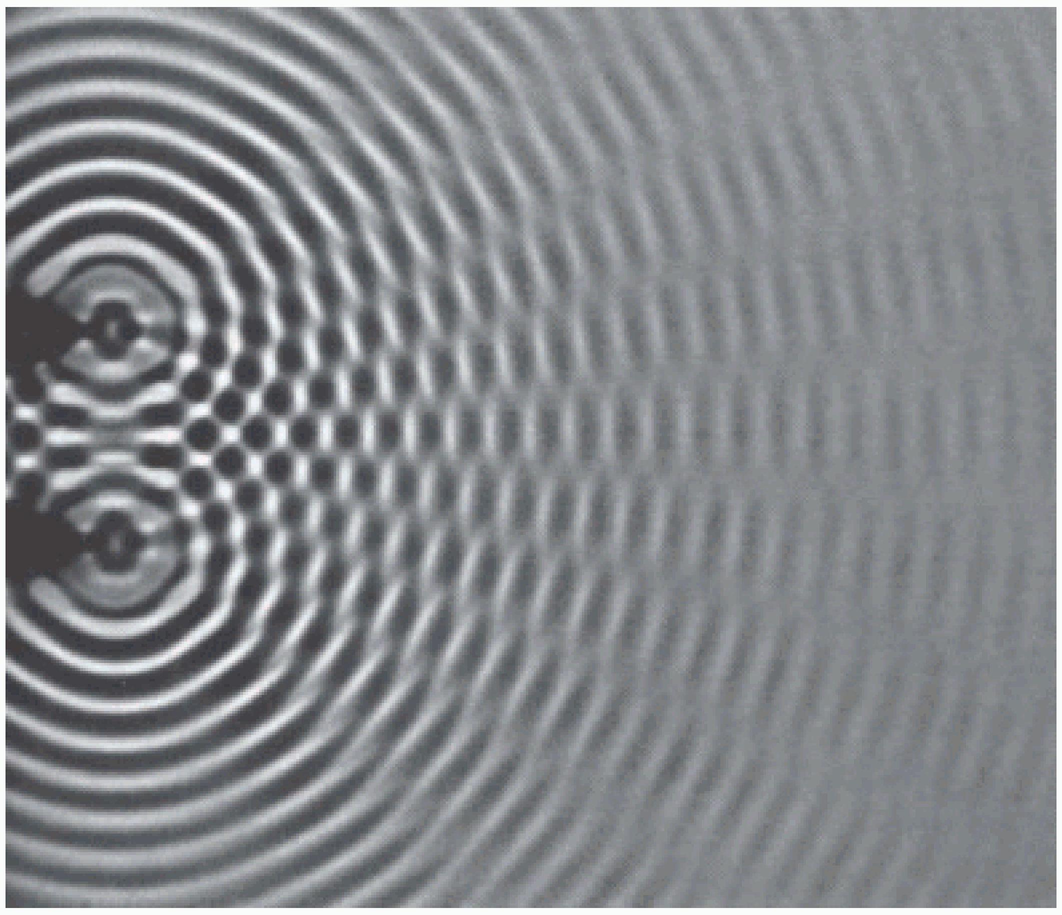


FIGURE 41-2. An interference pattern produced by water waves in a ripple tank. Two vibrating prongs at the left create two patterns of circular ripples, which overlap to give a pattern of maxima and minima in the waves. The maxima can be identified by the locations in the right half of the photograph where the shadows show the crests and valleys. The minima occur along lines where the shadows are less clearly defined.

41-2 DOUBLE-SLIT INTERFERENCE

Figure 41-3 illustrates our first method for dividing a light beam into two components that can then interfere with one another. A light beam, which we consider to be a train of plane waves such as might be obtained from a laser, is incident on a barrier into which two narrow slits have been cut. Portions of each incident wavefront pass through the slits, and so the slits can be considered as two sources of coherent light waves. (The spreading of light as it passes through the slits, illustrated in Fig. 41-3, is called *diffraction* and is discussed in the next chapter. A similar spreading occurs in the diffraction of water waves passing through a small opening, as we illustrated in Fig. 40-1. For now, we regard the slits as so narrow that each can be considered as a line of point sources of light, with each point source emitting spherical Huygens wavelets as discussed in Section 39-4.) Note that the two waves can overlap and interfere where they strike the screen. To simplify the analysis, we assume that the distance D between the slits and the screen is very much greater than the slit separation d. (Alternatively, we can place a lens between the slits and the screen to focus the emerging light on the screen, as we discuss later.)

When we view the screen, we see an alternating series of bright and dark bands, or *interference fringes,* corresponding respectively to maxima and minima in the intensity of

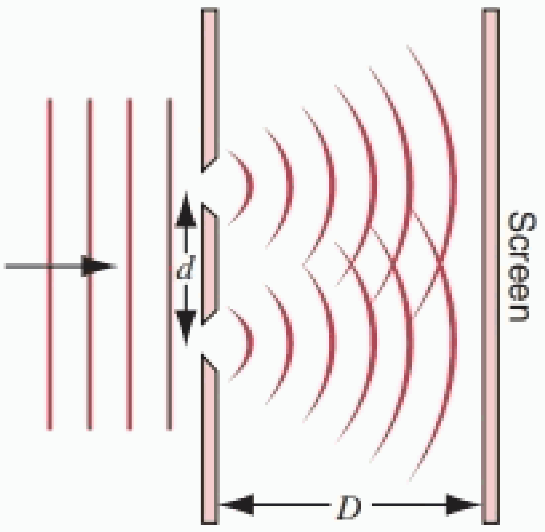


FIGURE 41-3. A train of plane light waves (for example, from a laser) is incident on a barrier into which are cut two narrow slits separated by a distance *d*. The widths of the slits are small compared with the wavelength, so that the waves passing through the slits spread out (diffract) and illuminate the screen.

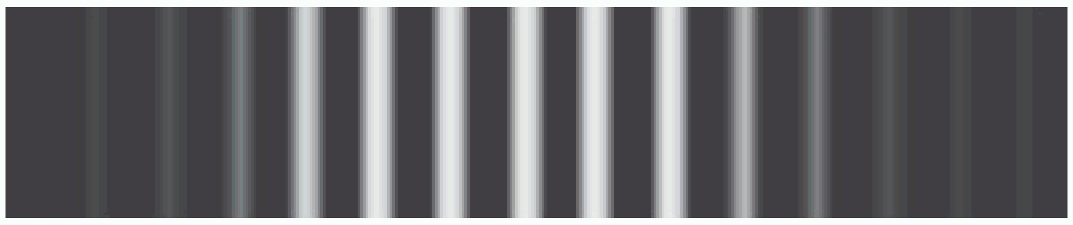


FIGURE 41-4. The interference pattern, consisting of bright and dark bands or fringes, that would appear on the screen of Fig. 41-3.

the light, as shown in Fig. 41-4. The right-hand edge of the ripple tank of Fig. 41-2 plays the role of the screen in Fig. 41-3. Along the "screen" in the ripple tank there is an alternating series of maxima and minima similar to Fig. 41-4.

To analyze the interference pattern, we consider waves from each slit that combine at an arbitrary point *P* on the screen *C* in Fig. 41-5. The point *P* is at distances of r_1 and r_2 from the narrow slits S_1 and S_2, respectively. The line S_2b is drawn so that the lines PS_2 and *Pb* have equal lengths. If *d*, the slit spacing, is much smaller than the distance *D* between the slits and the screen (the ratio d/D in the figure being exaggerated for clarity), S_2b is then almost perpendicular to both r_1 and r_2. This means that angle S_1S_2b is almost equal to angle *PaO,* both angles being marked θ in the figure; equivalently, the lines r_1 and r_2 may be taken as nearly parallel.

The two rays arriving at *P* in Fig. 41-5 from S_1 and S_2 were originally in phase at the source slits; both are derived from the same wavefront in the incident plane wave. Because the rays travel different distances (corresponding to different optical path lengths), they arrive at *P* with a phase difference. The number of wavelengths contained in the path difference S_1b determines the type of interference at *P*.

To have a *maximum* at *P,* the two rays must arrive in phase, and so S_1b $(= d \sin\theta)$ must contain a whole number of wavelengths (equivalent to shifting one of the waves in Fig. 41-1*a* by a whole number of cycles), or

$$S_1b = m\lambda \qquad m = 0, 1, 2, \ldots ,$$

which we can write as

$$d \sin\theta = m\lambda \qquad m = 0, \pm1, \pm2, \ldots \quad \text{(maxima).} \tag{41-1}$$

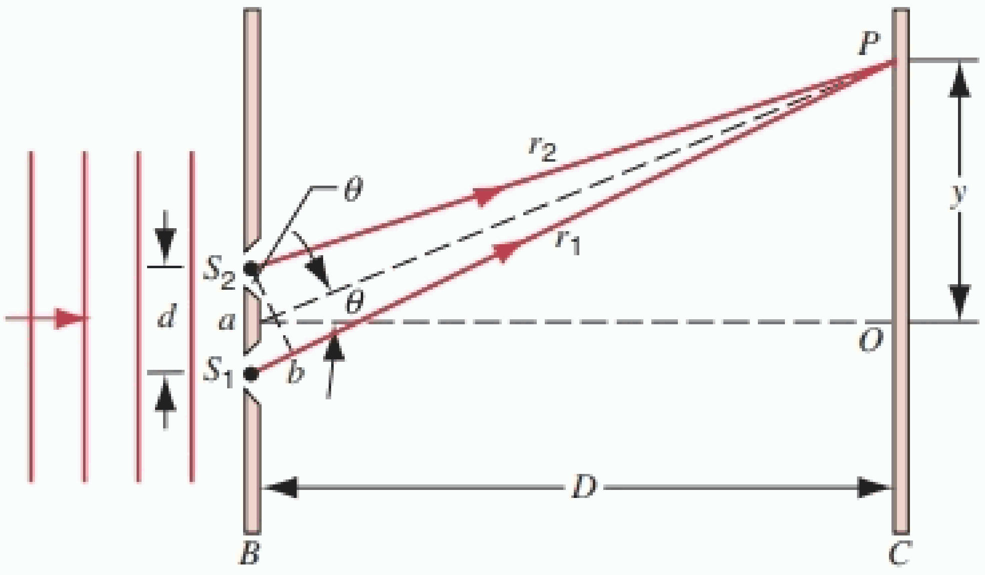


FIGURE 41-5. Rays from S_1 and S_2 combine at *P*. In actuality, $D \gg d$; the figure has been distorted for clarity. Point *a* is the midpoint between the slits.

Note that each maximum above *O* in Fig. 41-5 has a symmetrically located maximum below *O*; these correspond to using $m = -1, -2, \ldots$ in Eq. 41-1. The central maximum (at point *O*) is described by $m = 0$.

For a *minimum* at *P*, the two rays must differ in phase by an odd multiple of π (as in Fig. 41-1*b*), for which S_1b $(= d \sin\theta)$ must contain a half-integral number of wavelengths, or

$$d \sin\theta = (m + \tfrac{1}{2})\lambda \qquad m = 0, \pm1, \pm2, \ldots \quad \text{(minima).} \tag{41-2}$$

Negative values of *m* locate the minima on the lower half of the screen.

Often we are interested in the location of an interference maximum on the screen, indicated by the distance *y* measured from the center of the interference pattern in Fig. 41-5. It is convenient to make this calculation for the case in which *y* is small compared with the distance *D* between the slits and the screen. In this case the angle θ is small, and we can use the small-angle approximation $\sin\theta \approx \tan\theta$. With $\sin\theta = m\lambda/d$ (from Eq. 41-1) and $\tan\theta = y_m/D$ (from Fig. 41-5, where y_m represents the location of the *m*th maximum), we have

$$\frac{m\lambda}{d} = \frac{y_m}{D}$$

or

$$y_m = m\frac{\lambda D}{d} \qquad m = 0, \pm1, \pm2, \ldots \quad \text{(maxima).} \tag{41-3}$$

The separation between adjacent maxima is

$$\Delta y = y_{m+1} - y_m = (m+1)\frac{\lambda D}{d} - m\frac{\lambda D}{d}$$

or

$$\Delta y = \frac{\lambda D}{d}. \tag{41-4}$$

As long as θ is small, the separation of the maxima is independent of *m*; that is, the fringes are evenly spaced, as shown in Fig. 41-4. Furthermore, the spacing between adjacent minima is the same as the spacing between the maxima.

It is mathematically simpler to deal with plane waves that are incident on the double slit and that emerge from it. However, plane wavefronts do not form an image on a screen at any finite distance *D* from the slits. We therefore often use a lens, as shown in Fig. 41-6, to focus parallel rays from the slits onto the screen. Light focused at *P* must have struck the lens parallel to the line *Px* drawn from *P* through the center of the lens. Under these conditions, rays r_1 and r_2 are strictly parallel, even though the requirement $D \gg d$ may not be met.

If a lens is used between the slits and the screen, it may seem that a phase difference should develop between the rays beyond the plane represented by S_2b, the geometrical path lengths between this plane and *P* being clearly

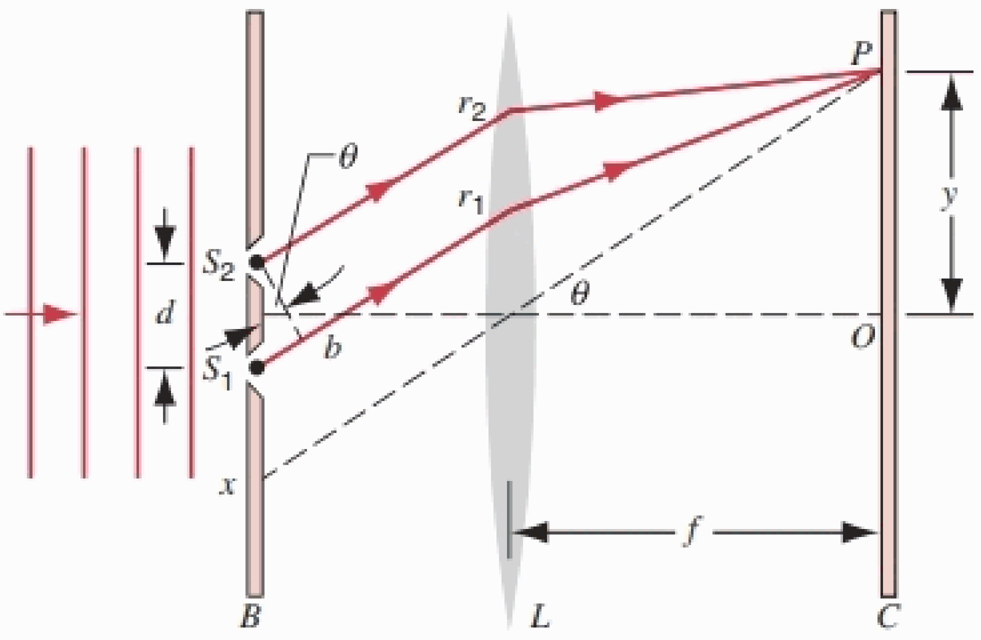


FIGURE 41-6. A lens is used to produce the interference fringes. Compare with Fig. 41-5. In actuality, $f \gg d$; the figure is again distorted for clarity.

different. In Section 40-5, however, we saw that for parallel rays focused by a lens the *optical* path lengths are equal. Two rays with equal optical path lengths contain the same number of wavelengths, so no phase difference occurs as a result of the light passing through the lens.

Young's Double-Slit Experiment

An interference experiment of the kind described above was first performed in 1801 by Thomas Young.* Young's experiment provided the first conclusive proof of the wave nature of light. Because, as indicated by Eq. 41-3, the locations of the interference maxima depend on the wavelength, Young's experiments provided the first direct measurement of the wavelength of light.

There were of course no lasers in Young's time, so he created a source of coherent light by allowing sunlight to fall on a narrow opening S_0, as shown in Fig. 41-7. The spreading wavelets from S_0 give coherent wavefronts that pass through the two openings. Young used pinholes rather than slits for his experiments, and as a result the interference pattern was more complicated than that of Fig. 41-4. Nevertheless, his conclusions regarding the wave nature of light were unambiguous. Even when carried out using a laser, the double-slit experiment is often known as Young's experiment.

SAMPLE PROBLEM 41-1. The double-slit arrangement in Fig. 41-5 is illuminated with light from a mercury vapor lamp filtered so that only the strong green line ($\lambda = 546$ nm) is visible. The slits are 0.12 mm apart, and the screen on which the interference pattern appears is 55 cm away. (*a*) What is the angular position of the first minimum? Of the tenth maximum? (*b*) What is the distance on the screen between the adjacent maxima?

Solution (*a*) At the first minimum we put $m = 0$ in Eq. 41-2,

$$\sin\theta = \frac{(m+\frac{1}{2})\lambda}{d} = \frac{(\frac{1}{2})(546\times10^{-9}\text{ m})}{0.12\times10^{-3}\text{ m}} = 0.0023.$$

This value for $\sin\theta$ is so small that we can take it to be the value of θ, expressed in radians; expressed in degrees it is 0.13°.

At the tenth maximum (not counting the central maximum) we must put $m = 10$ in Eq. 41-1. Doing so and calculating as before we find an angular position of 2.6°. For these conditions we see that the angular spread of the first dozen or so fringes is small.

(*b*) From Eq. 41-4, we have

$$\Delta y = \frac{\lambda D}{d} = \frac{(546\times10^{-9}\text{ m})(55\times10^{-2}\text{ m})}{0.12\times10^{-3}\text{ m}} = 2.5\text{ mm}.$$

41-3 COHERENCE

For an interference pattern to occur at all, *the phase difference at points on the screen must not change with time.* If this occurs, we say that the beams from slits S_1 and S_2 are completely *coherent.* Coherence can occur for any type of waves. For example, coherent sound waves can be obtained by driving two different loudspeakers from the same audio oscillator, and coherent radio waves similarly result when two different antennas are connected to the same electromagnetic oscillator.

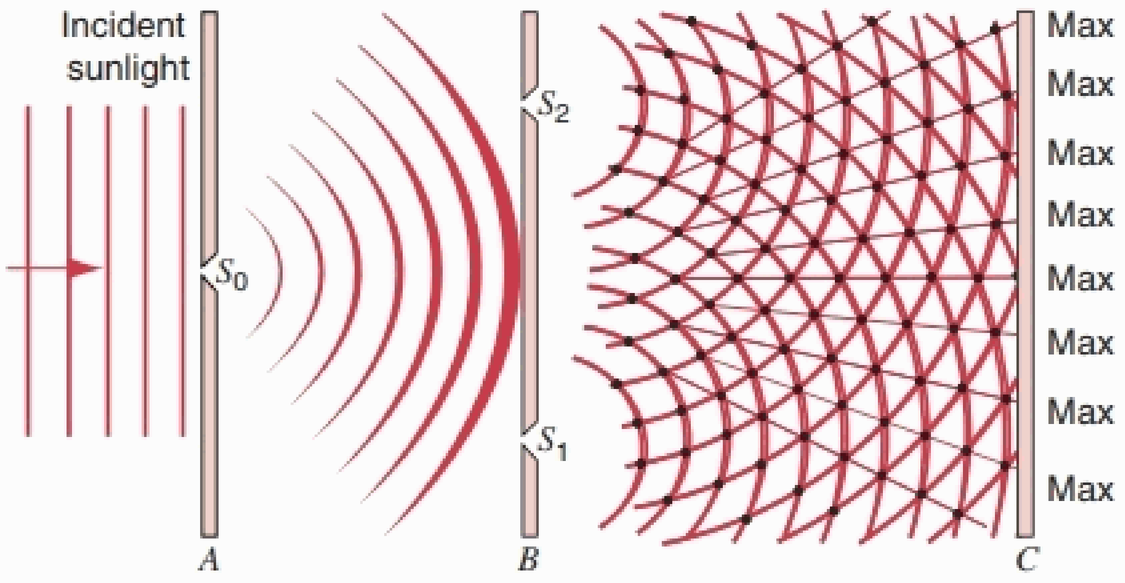


FIGURE 41-7. In Young's interference experiment, sunlight diffracted from pinhole S_0 falls on pinholes S_1 and S_2 in screen *B*. Light diffracted from these two pinholes overlaps on screen *C*, producing the interference pattern.

---

*Thomas Young (1773–1829) was originally trained as a physician. His interest in sense perception and vision led him to physics and the study of light. Among his other scientific accomplishments were studies of surface tension and elasticity, which was recognized with the naming of the elastic modulus, now known as Young's modulus, in his honor. He was also noted for his interest in hieroglyphics and contributed to the deciphering of the Rosetta stone, which provided the first understanding of ancient Egyptian languages.

Suppose instead that slits S_1 and S_2 are replaced by two completely independent light sources, such as two fine incandescent wires placed side by side. No interference fringes appear on screen C; there is only a relatively uniform illumination. We can interpret this if we make the reasonable assumption that for completely independent light sources the phase difference between the two beams arriving at any point on the screen varies with time in a random way. At a certain instant, conditions may be right for cancellation, and a short time later (perhaps 10^{-8} s) they may be right for reinforcement. The eye cannot follow these rapid variations and sees only a uniform illumination. The intensity at any point is equal to the sum of the intensities that each source S_1 and S_2 produces separately at that point. Under these conditions the two beams emerging from S_1 and S_2 are said to be completely *incoherent.*

To find the intensity resulting from the overlap of completely *coherent* light beams, we (1) add the wave amplitudes, taking the (constant) phase difference properly into account, and then (2) square the resultant amplitude to obtain a quantity proportional to the resultant intensity. For completely *incoherent* light beams, on the other hand, we (1) square the individual amplitudes to obtain quantities proportional to the individual intensities and then (2) add the individual intensities to obtain the resultant intensity. These procedures agree with the experimental facts that for completely independent light sources the resultant intensity at every point is always greater than the intensity produced at that point by either light source acting alone, but for coherent sources the intensity at some points may be less than that produced by either source alone.

How can we produce coherent waves? A plane electromagnetic wave produced by a cavity oscillator (see Section 38-3) connected to an antenna produces a wave that we can represent (at large distances from the antenna) in the form of Fig. 41-8*a*. This wave has essentially infinite extent in time, including both future times ($t > 0$) and past times ($t < 0$). At any point, as the wave passes by, the wave disturbance ($\vec{\mathbf{E}}$ or $\vec{\mathbf{B}}$) varies with time in a perfectly periodic way. Moreover, at any instant in *time* the wave disturbance varies with *distance* in a perfectly periodic way. If two identical oscillators produce waves that travel along the same direction, the phase difference between them at one location along the direction of propagation would be the same at all other locations along that direction, and the phase difference at a given location has the same value at all times. For another example, the two vibrating prongs of the ripple tank of Fig. 41-2 generate two coherent waves in the water.

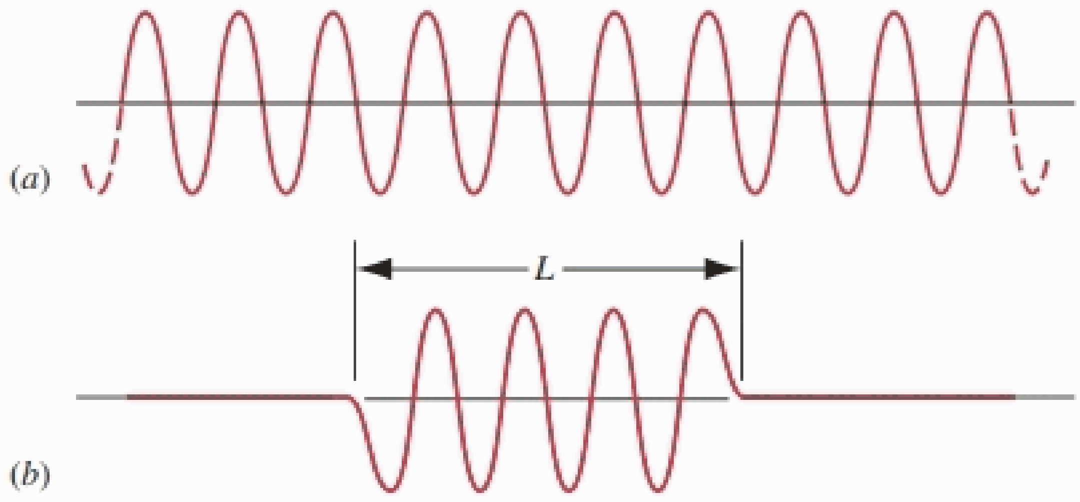


FIGURE 41-8. (*a*) A section of an infinite wave. (*b*) A wavetrain of finite length L.

If we turn instead to common sources of visible light, such as incandescent wires or an electric discharge passing through a gas, we become aware of a fundamental difference. In both of these sources the fundamental light emission processes occur in individual atoms, and these atoms do not act together in a cooperative (that is, *coherent*) way. The act of light emission by a single atom takes, in a typical case, about 10^{-8} s, and the emitted light is properly described as a finite *wavetrain* (Fig. 41-8*b*) rather than as an infinite wave (Fig. 41-8*a*). For emission times such as these, the wavetrains are a few meters long. For actual light sources, such as low-pressure gas discharge tubes, the wavetrains are typically of the order of centimeters long. This is the limit of distances over which light from such sources remains coherent.

Interference effects from ordinary light sources can be produced by putting a very narrow slit (such as S_0 in Fig. 41-7) directly in front of the source. This ensures that the wavetrains that strike slits S_1 and S_2 in screen B in Fig. 41-7 originate from the same small region of the source. If the path lengths from all points within slit S_0 to S_1 and S_2 are nearly equal, light passing through the double slits is in phase, and a stationary interference pattern is produced on the screen C. If slit S_0 is made so wide that there are points within S_0 for which the path lengths to S_1 and S_2 differ by one-half wavelength, the light passing through the double slits from sources at such points would be out of phase; this light contributes maxima on the screen where there were previously minima and minima where there were previously maxima. The effect on the screen becomes an incoherent superposition of the light from effectively many sources, thereby washing out the interference pattern.

If the slit S_0 is small enough, a given constant phase difference is maintained at any point on the screen C between beams passing through the slits S_1 and S_2. We can regard this light as coherent, within a distance characterized by the length of its wavetrain.

If the width of slit S_0 in Fig. 41-7 is gradually increased, it is observed experimentally that the maxima of the interference fringes become reduced in intensity and that the intensity in the fringe minima is no longer strictly zero. In other words, the fringes become less distinct. If S_0 is opened extremely wide, the lowering of the maximum intensity and the raising of the minimum intensity are such that the fringes disappear, leaving only a uniform illumination characteristic of two completely incoherent sources. As the width of slit S_0 increases, the beams from S_1 and S_2 pass continuously from a condition of complete coherence to one of complete incoherence. When not at either of these two limits, the beams are said to be partially coherent.

In contrast to the atoms of the filament of an incandescent lightbulb, the atoms in a laser emit their light in a completely cooperative way (see Section 48-9). Laser light is therefore highly coherent, and we can observe interference simply by illuminating the double slit with a laser beam, as in Fig. 41-3, without the necessity of using the slit S_0. The laser beam is not perfectly coherent; the wavetrains of a laser may have lengths of tens to hundreds of kilometers, meaning that we can split and recombine a laser beam and observe interference effects over large distances. One application of this coherence is in holography (see Section 43-5), in which one of the split beams is reflected from an object and the interference pattern between the direct and reflected beams is stored on photographic film; a three-dimensional image of the object can be reconstructed from the film. Other applications that depend on the long coherence length of lasers include communication over long distances using optical signals and interference measurements that track the movement of tectonic plates over the Earth's surface.

41-4 INTENSITY IN DOUBLE-SLIT INTERFERENCE

Equations 41-1 and 41-2 give the locations of the maxima and minima of the interference pattern. They do not, however, indicate how the intensity varies between the maxima and minima. In this section we derive an expression for the intensity I at any point P located by the angle θ in Fig. 41-5.

Let us assume that the electric field components* of the two waves in Fig. 41-5 vary with time at point P as

$$E_1 = E_0 \sin \omega t \tag{41-5}$$

and

$$E_2 = E_0 \sin (\omega t + \phi), \tag{41-6}$$

where $\omega\ (=2\pi f)$ is the angular frequency of the waves and ϕ is the phase difference between them. Note that ϕ depends on the location of point P, which is described by the angle θ in Fig. 41-5. We assume that the slits are so narrow that the diffracted light from each slit illuminates the central portion of the screen uniformly (Fig. 41-3). This means that near the center of the screen E_0 is independent of the position of P, that is, of the value of θ.

If the slit separation d is much smaller than the distance D to the screen, the E vectors from the two interfering waves are nearly parallel, and we can replace the vector sum of the E fields with the sum of their components,

$$E = E_1 + E_2, \tag{41-7}$$

which, as we prove later in this section, can be written

$$E = E_\theta \sin (\omega t + \beta), \tag{41-8}$$

where the phase angle β is

$$\beta = \tfrac{1}{2}\phi \tag{41-9}$$

and the amplitude is

$$E_\theta = 2E_0 \cos \beta. \tag{41-10}$$

The amplitude E_θ of the resultant wave disturbance, which determines the intensity of the interference fringes, depends on β, which in turn depends on the value of θ, that is, on the location of point P in Fig. 41-5. The maximum possible value of the amplitude E_θ is $2E_0$, equal to twice the amplitude E_0 of the combining waves, corresponding to complete reinforcement.

In Section 38-6, we showed that the intensity I of an electromagnetic wave is proportional to the square of its electric field amplitude E_m (see Eq. 38-26):

$$I = \frac{1}{2\mu_0 c} E_m^2. \tag{41-11}$$

The ratio of the intensities of two light waves can therefore be expressed as the ratio of the squares of the amplitudes of their electric fields. If I_θ is the intensity of the resultant wave at P, and I_0 is the intensity that each single wave acting alone would produce, then

$$\frac{I_\theta}{I_0} = \left(\frac{E_\theta}{E_0}\right)^2. \tag{41-12}$$

Combining Eqs. 41-10 and 41-12, we obtain

$$I_\theta = 4I_0 \cos^2 \beta. \tag{41-13}$$

Note that the intensity of the resultant wave at any point P varies from zero [for a point at which $\phi\ (=2\beta) = \pi$, say] to four times the intensity I_0 of each individual wave [for a point at which $\phi\ (=2\beta) = 0$, say]. Let us compute I_θ as a function of the angle θ in Fig. 41-5.

The phase difference ϕ in Eq. 41-6 is associated with the path difference S_1b in Fig. 41-5. If S_1b is $\frac{1}{2}\lambda$, ϕ is π; if S_1b is λ, ϕ is 2π, and so forth. In general,

$$\frac{\text{phase difference}}{2\pi} = \frac{\text{path difference}}{\lambda}. \tag{41-14}$$

Letting ϕ be the phase difference and recalling that the path difference in Fig. 41-5 is $d \sin \theta$, we can write this as

$$\phi = \frac{2\pi}{\lambda}(d \sin \theta),$$

or, using Eq. 41-9,

$$\beta = \tfrac{1}{2}\phi = \frac{\pi d}{\lambda} \sin \theta. \tag{41-15}$$

The intensity at any θ can therefore be written

$$I_\theta = 4I_0 \cos^2 \tfrac{1}{2}\phi \tag{41-16}$$

or

$$I_\theta = 4I_0 \cos^2 \left(\frac{\pi d \sin \theta}{\lambda}\right). \tag{41-17}$$

---

*We could characterize the light wave either by its electric field $\vec{\mathbf{E}}$ or its magnetic field $\vec{\mathbf{B}}$. We generally use $\vec{\mathbf{E}}$ rather than $\vec{\mathbf{B}}$, because the effects of $\vec{\mathbf{B}}$ on the human eye and on various light detectors are exceedingly small.

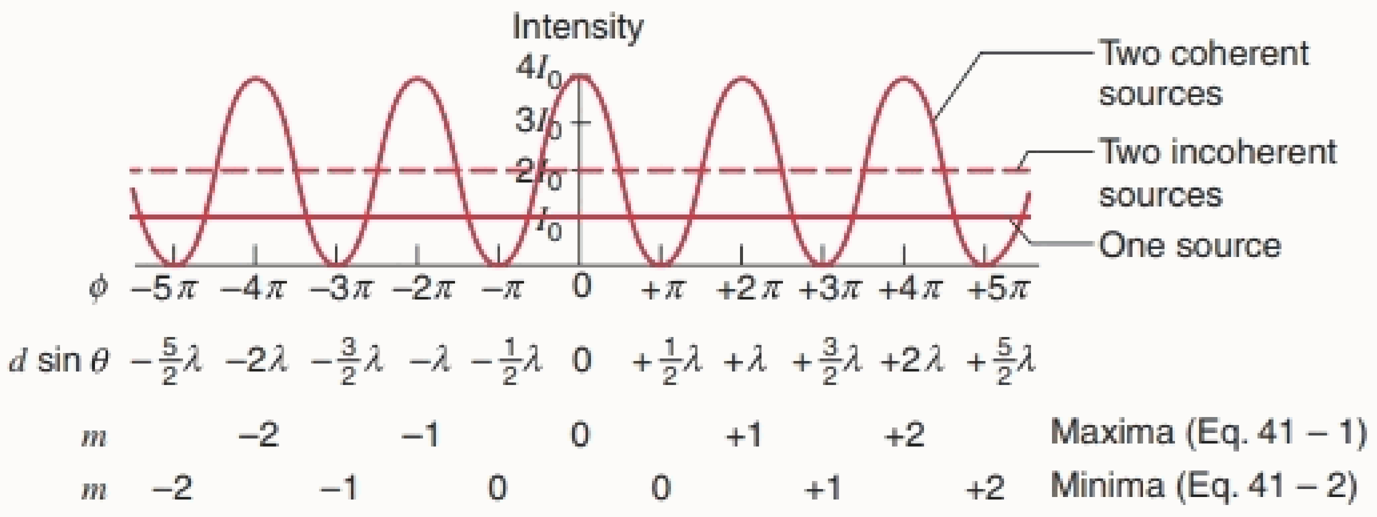


FIGURE 41-9. The intensity pattern for double-slit interference, assuming that the two interfering waves illuminate this region of the screen uniformly; that is, I_0 is independent of position.

From Eq. 41-16, we see that intensity maxima occur where $\cos^2 \frac{1}{2}\phi = 1$, or

$$\phi = 2m\pi \qquad m = 0, \pm 1, \pm 2, \ldots .$$

Using Eq. 41-15, we can write this as

$$d \sin\theta = m\lambda \qquad m = 0, \pm 1, \pm 2, \ldots \qquad \text{(maxima)},$$

which is the same as Eq. 41-1. Intensity minima occur, according to Eq. 41-16, where $\cos^2 \frac{1}{2}\phi = 0$, or

$$\phi = (2m + 1)\pi \qquad m = 0, \pm 1, \pm 2, \ldots ,$$

which we write using Eq. 41-15 as

$$d \sin\theta = (m + \tfrac{1}{2})\lambda \qquad m = 0, \pm 1, \pm 2, \ldots \qquad \text{(minima)},$$

in agreement with Eq. 41-2.

Figure 41-9 shows the intensity pattern for double-slit interference. The horizontal solid line is I_0; this describes the (uniform) intensity pattern on the screen if one of the slits is covered up. If the two sources were incoherent, the intensity would be uniform over the screen and would be $2I_0$, indicated by the horizontal dashed line of Fig. 41-9. For coherent sources we expect the energy to be merely redistributed over the screen, because energy is neither created nor destroyed by the interference process. Thus the *average* intensity in the interference pattern should be $2I_0$, as for incoherent sources. This follows at once if, in Eq. 41-16, we substitute for the cosine-squared term the value $\frac{1}{2}$, which always results when we average the square of a sine or a cosine term over one or more half-cycles.

Adding Wave Disturbances

We now derive Eqs. 41-8 to 41-10 for the combined electric field of the light in double-slit interference. This derivation can be done algebraically, using the methods of Section 18-8. However, the algebraic method becomes extremely difficult when we wish to add more than two wave disturbances, as we do in later chapters. We therefore use a graphical method, which proves to be convenient in more complicated situations. This method is based on rotating *phasors* and is similar to that used in the analysis of alternating current circuits in Chapter 37.

A sinusoidal wave disturbance such as that of Eq. 41-5 can be represented graphically using a rotating phasor. In Fig. 41-10*a*, a phasor of magnitude E_0 is allowed to rotate about the origin in a counterclockwise direction with an angular frequency ω. The varying wave disturbance E_1 ($= E_0 \sin \omega t$, as in Eq. 41-5) is represented by the projection of this phasor on the vertical axis.

A second wave disturbance E_2, given by Eq. 41-6, which has the same amplitude E_0 but a phase difference ϕ with respect to E_1, can be represented graphically (Fig. 41-10*b*) as the projection on the vertical axis of a second phasor of the same magnitude E_0, which makes an angle ϕ with the first phasor. Thus $E_2 = E_0 \sin(\omega t + \phi)$, as in Eq. 41-6.

The sum E of E_1 and E_2 is the sum of the projections of the two phasors on the vertical axis. This is revealed more clearly if we redraw the phasors, as in Fig. 41-10*c*, placing the tail of one arrow at the head of the other, maintaining the proper phase difference, and letting the whole assembly rotate counterclockwise about the origin.

In Fig. 41-10*c*, E can also be regarded as the projection on the vertical axis of a phasor of length E_θ, which is the vector sum of the two phasors of magnitude E_0. From that figure, we see that the projection can be written

$$E = E_\theta \sin(\omega t + \beta),$$

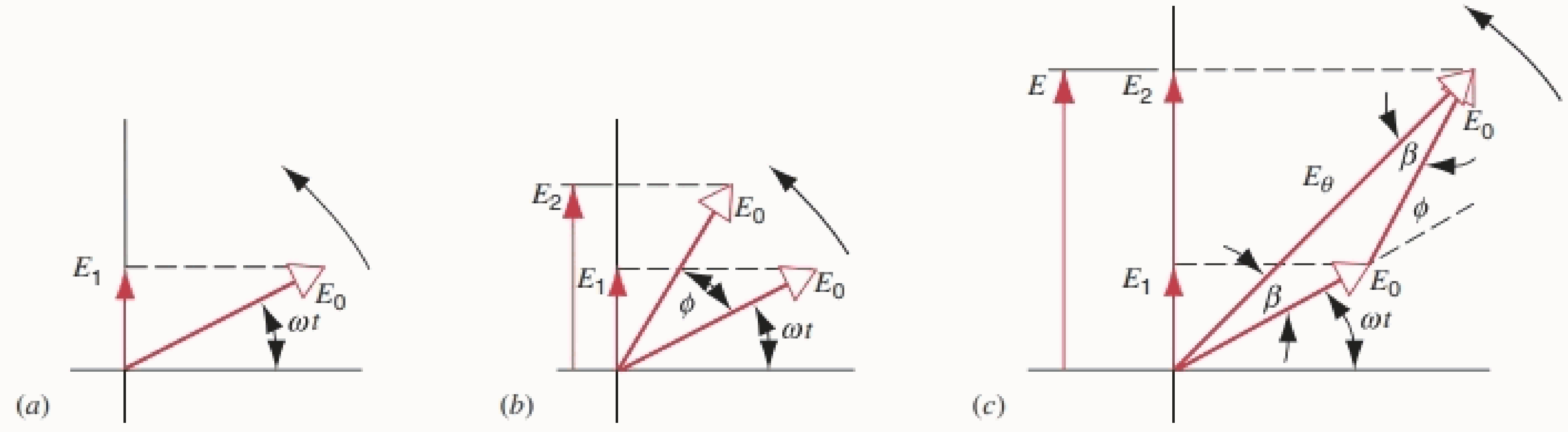


FIGURE 41-10. (*a*) A time-varying wave E_1 is represented by a rotating vector or phasor. (*b*) Two waves E_1 and E_2 differing in phase by ϕ. (*c*) Another way of drawing (*b*).

in agreement with Eq. 41-8. Note that the (algebraic) sum of the projections of the two phasors is equal to the projection of the (vector) sum of the two phasors.

In most problems in optics we are concerned only with the *amplitude* E_θ of the resultant wave disturbance and not with its time variation. This is because the eye and other common measuring instruments respond to the resultant intensity of the light (that is, to the square of the amplitude) and cannot detect the rapid time variations that characterize visible light. For sodium light ($\lambda = 589$ nm), for example, the frequency $f\,(=\omega/2\pi)$ is 5.1×10^{14} Hz. Often, then, we need not consider the rotation of the phasors but can confine our attention to finding the amplitude of the resultant phasor.

In Fig. 41-10*c* the three phasors form an isosceles triangle whose sides have lengths E_0, E_0, and E_θ. In any triangle, an exterior angle (ϕ in this case) is equal to the sum of the two opposite interior angles (β and β), and so

$$\beta = \tfrac{1}{2}\phi.$$

It is also clear from Fig. 41-10*c* that the length of the base of this triangle is

$$E_\theta = 2E_0 \cos \beta.$$

These results are identical with Eqs. 41-9 and 41-10.

In a more general case we might want to find the resultant of more than two sinusoidally varying wave disturbances. The general procedure is the following:

1. Construct a series of phasors representing the functions to be added. Draw them end to end, maintaining the proper phase relationships between adjacent phasors.

2. Construct the sum of this phasor array, analogous to a sum of vectors. The length of the resulting phasor gives the amplitude of the electric field. The angle between it and the first phasor is the phase of the resultant with respect to this first phasor. The projection of this phasor on the vertical axis gives the time variation of the resultant wave disturbance.

SAMPLE PROBLEM 41-2. Find graphically the resultant $E(t)$ of the following wave disturbances:

$$E_1 = E_0 \sin \omega t,$$

$$E_2 = E_0 \sin (\omega t + 15^\circ),$$

$$E_3 = E_0 \sin (\omega t + 30^\circ),$$

$$E_4 = E_0 \sin (\omega t + 45^\circ).$$

Solution Figure 41-11 shows the assembly of four phasors that represent these functions. The phase angle ϕ between successive phasors is 15°.

In any closed n-sided polygon, the sum of the interior angles is $(n-2)\pi$, so in the five-sided figure *abcdea,* the sum of the interior angles is 540°. The interior angles at b, c, and d are each $180^\circ - 15^\circ = 165^\circ$, from which we deduce

$$\beta = \tfrac{1}{2}[540^\circ - 3(165^\circ)] = 22.5^\circ.$$

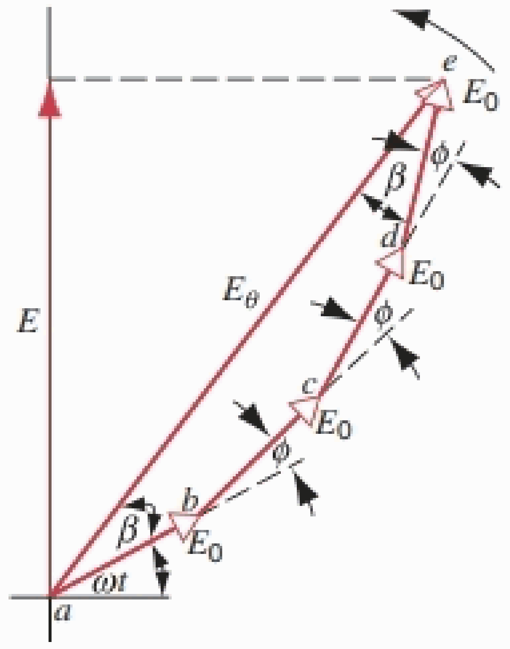


FIGURE 41-11. Sample Problem 41-2. Four waves are added graphically, using the method of phasors.

The length of side *ae* can be written

$$ae = ab \cos \beta + bc \cos (\beta - 15^\circ) + cd \cos (\beta - 15^\circ) + de \cos \beta.$$

With $ae = E_\theta$ and $ab = bc = cd = de = E_0$, we deduce

$$E_\theta = 3.83\,E_0.$$

The resultant $E(t)$ is then the projection of E_θ on the vertical axis:

$$E(t) = E_\theta \sin (\omega t + \beta) = 3.83\,E_0 \sin (\omega t + 22.5^\circ).$$

41-5 INTERFERENCE FROM THIN FILMS

The colors that we see when sunlight falls on a soap bubble, an oil slick, or a ruby-throated hummingbird are caused by the interference of light waves reflected from the front and back surfaces of thin, transparent films. The film thickness is typically of the order of magnitude of the wavelength of light. Thin films deposited on optical components, such as camera lenses, can reduce reflection and enhance the intensity of the transmitted light. Thin coatings on windows can enhance the reflectivity for infrared radiation while having less effect on the visible radiation. In this way it is possible to reduce the heating effect of sunlight on a building.

Depending on its thickness, a thin film can be perfectly reflecting or perfectly transmitting for light of a given wavelength, as shown in Fig. 41-12. These effects result from constructive or destructive interference.

Figure 41-13 shows a transparent film of uniform thickness d illuminated by monochromatic light of wavelength λ from a point source S. The eye is positioned so that a particular incident ray i from the source enters the eye as ray r_1 after reflection from the front surface of the film at a. The incident ray also enters the film at a as a refracted ray and is reflected from the back surface of the film at b; it then emerges from the front surface of the film at c and also enters the eye, as ray r_2. Rays r_1 and r_2 are parallel in the

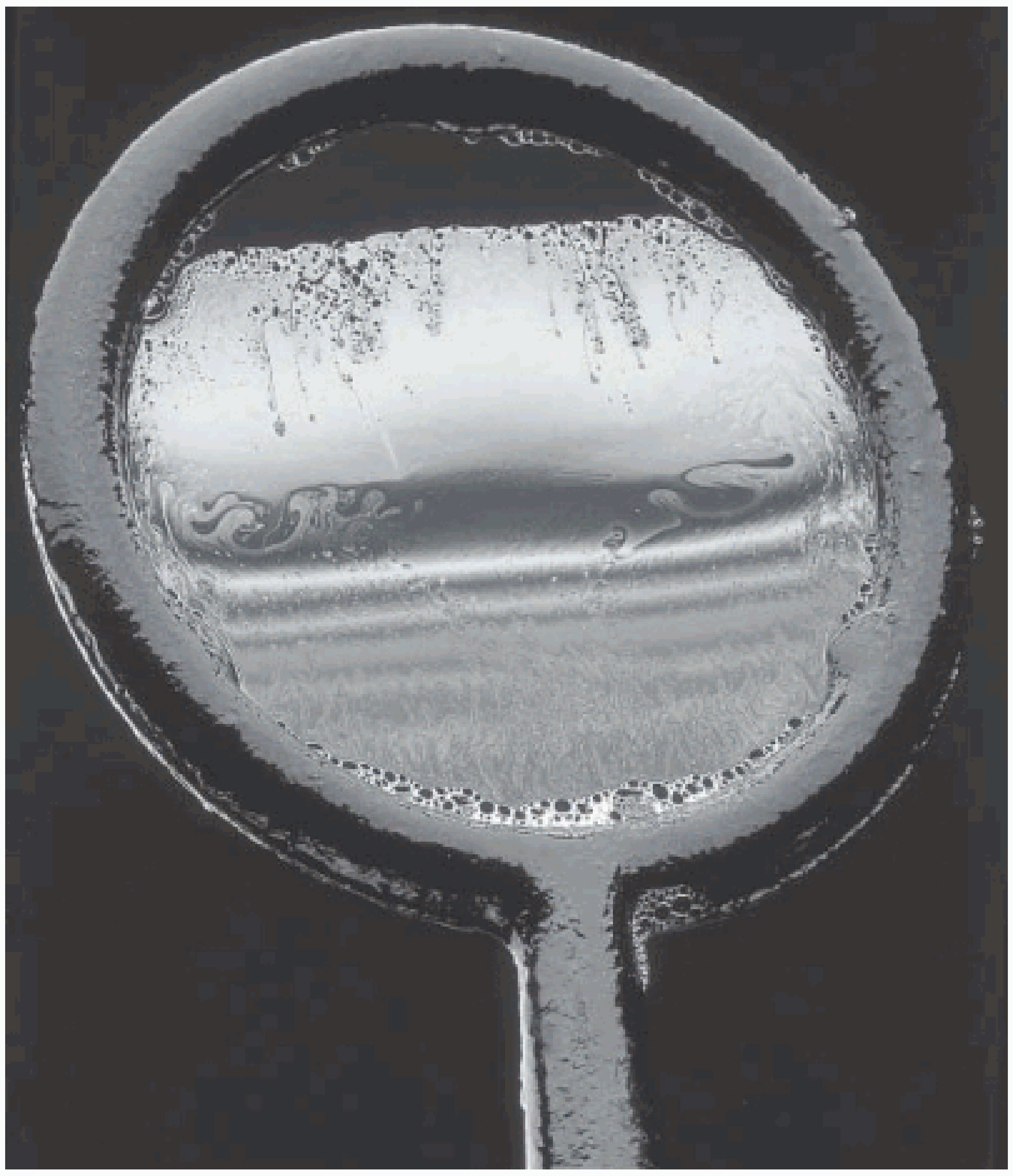


FIGURE 41-12. A soapy water film on a vertical loop, viewed by reflected light. The black segment at the top is not a torn film. It occurs because the film, by drainage, is so thin there that destructive interference occurs between light reflected from the front and back surfaces of the film.

geometry of Fig. 41-13. Having originated at the same point source, they are also coherent and so are capable of interfering. Because these two rays have traveled over paths of different lengths, have traversed different media, and—as we shall see—have suffered different kinds of reflections at a and b, there is a phase difference between them. The intensity perceived by the eye, as the parallel rays from the region ac of the film enter it, is determined by this phase difference.

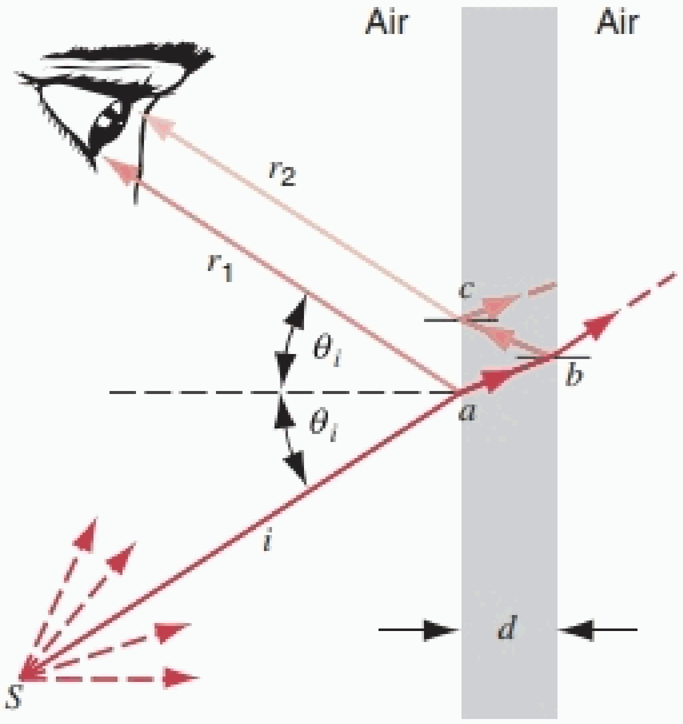


FIGURE 41-13. A thin film is viewed by light reflected from a source S. Waves reflected from the front and back surfaces enter the eye as shown, and the intensity of the resultant light wave is determined by the phase difference between the combining waves. The medium on either side of the film is assumed to be air.

There are two contributions to this phase difference. It is apparent from Fig. 41-13 that ray r_2 travels a longer path than ray r_1, because r_2 must pass twice through the film before reaching the eye. This additional path contributes to the phase difference between the rays. Another contribution to the path difference is a sudden change in phase of 180° that can occur when a wave is reflected. Before we can discuss the interference of r_1 and r_2, we must consider this second contribution to the phase change in detail.

Phase Change on Reflection

Let us consider a very thin film, in which $2d$ is much less than one wavelength. Since the additional path that r_2 must travel is about $2d$, the contribution to the phase difference from the passage of r_2 twice through the film would be close to zero. If the phase difference between r_1 and r_2 were due only to this additional path, then the total phase difference should be close to zero, in which case r_1 and r_2 would interfere constructively. We would therefore expect the film to look bright in reflected light. However, Fig. 41-12 clearly shows that this is not the case. The top portion of the film, where it is thinnest because of the action of gravity, appears dark rather than bright. It is apparent that *destructive* rather than constructive interference occurs at the thinnest portion of the film.

To explain this and many similar phenomena, we conclude that one or the other of the two rays of Fig. 41-13 must suffer an abrupt phase change of π (= 180°) when it is reflected at the air–film interface. As it turns out, only the ray reflected from the front surface suffers this phase change. The other ray is not changed abruptly in phase, either on transmission through the front surface or on reflection at the back surface.

In Section 18-9 we discussed phase changes on reflection for transverse waves in strings. To extend these ideas, consider the composite string of Fig. 41-14, which consists of two parts with different masses per unit length, stretched to a given tension. A pulse in the heavier string moves to the

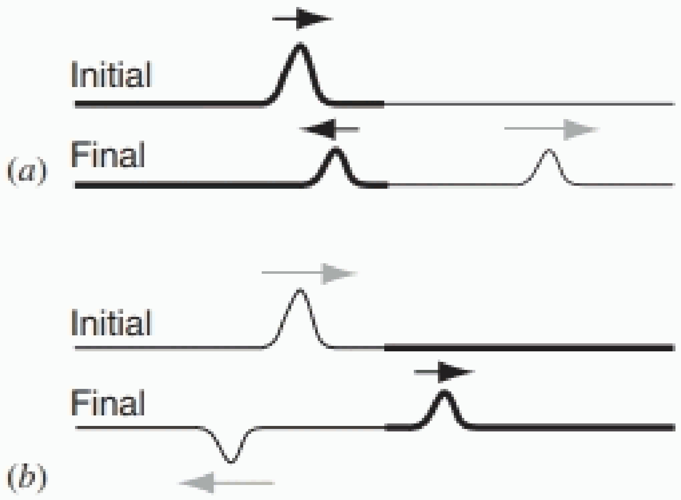


FIGURE 41-14. Phase changes on reflection at a junction between two strings of different linear mass densities. The wave speed is greater in the lighter string. (*a*) The incident pulse is in the heavier string. (*b*) The incident pulse is in the lighter string.

right in Fig. 41-14*a*, approaching the junction. Later there will be a reflected pulse (moving back to the left) and a transmitted pulse (continuing to the right). In this case the reflected pulse is *in phase* with the incident pulse. In Fig. 41-14*b* the situation is reversed, the incident pulse now being in the lighter string. In this case the reflected pulse differs in phase from the incident pulse by π (= 180°). In each case the transmitted pulse is in phase with the incident pulse.

Figure 41-14*a* suggests a light wave in glass, say, approaching a surface beyond which there is a less optically dense medium (one of lower index of refraction) such as air. Figure 41-14*b* suggests a light wave in air approaching glass. We can therefore sum up the optical situation for reflection at an interface as follows:

> *If the medium beyond the interface has a higher index of refraction, the reflected wave undergoes a phase change of π (180°). If the medium beyond the interface has a lower index of refraction, there is no phase change of the reflected wave.**

In the case of Fig. 41-13, a phase change of π occurs for the reflection of ray r_1 at point a, because the medium beyond the interface (the film) has a higher index of refraction than the medium in which the ray is traveling (air). No phase change occurs for the reflection of ray r_2 at point b because the ray is traveling in the film, and the air beyond the interface has a *lower* index of refraction than the film. No phase change occurs for the refraction of ray r_2 at point c, because there is never a phase change on refraction; similarly, the ray that is refracted into the film at point a experiences no phase change.

Equations for Thin-Film Interference

It is important to note that the wavelength that determines the type of interference is the wavelength λ_n in the film, not the wavelength λ in air. We can also say that we are interested in *optical path differences* and not in *geometrical path differences*. The wavelengths λ_n and λ are related by Eq. 39-14:

$$\lambda_n = \lambda/n, \tag{41-18}$$

where n is the index of refraction of the film.

Whether rays r_1 and r_2 interfere constructively or destructively depends on the phase difference between them. If the phase difference is $0, 2\pi, 4\pi, \ldots$, then the interference is constructive. Equivalently, we can say that the path difference for constructive interference is $0, \lambda_n, 2\lambda_n, \ldots$, or simply $m\lambda_n$ with $m = 0, 1, 2, \ldots$. For destructive interference, the phase difference is $\pi, 3\pi, 5\pi, \ldots$, and the equivalent path difference is $\lambda_n/2, 3\lambda_n/2, 5\lambda_n/2, \ldots$, or $(m + \frac{1}{2})\lambda_n$ with $m = 0, 1, 2, \ldots$.

We will simplify the problem somewhat by assuming near-normal incidence; that is, $\theta_i \approx 0$ in Fig. 41-13. In this case the contribution to the path difference due to the travel of ray r_2 through the film is approximately $2d$. The other possible contributions to the total path difference between r_1 and r_2 correspond to the phase difference of π (or path difference of one-half wavelength) that might occur on reflection at the front surface of the film (if the index of refraction of the film is greater than the index of refraction of the medium from which the ray originates), and the phase difference of π (or path difference of one-half wavelength) that might occur upon reflection at the back surface of the film (if the index of refraction of the film is smaller than the index of refraction of the medium beyond the back surface). We can then write an expression for the total path difference between rays r_1 and r_2 as follows:

$$\text{path difference} = 2d + \underset{\substack{\uparrow \\ \text{front} \\ \text{surface}}}{\overset{?}{\tfrac{1}{2}\lambda_n}} + \underset{\substack{\uparrow \\ \text{back} \\ \text{surface}}}{\overset{?}{\tfrac{1}{2}\lambda_n}}. \tag{41-19}$$

The question marks are meant as a reminder that, depending on the relative index of refraction of the film in comparison with what is on either side of the film, we might need to include neither of the extra terms, or perhaps one of them, or perhaps both of them.

Let us consider again the example of Fig. 41-13, in which a film of index of refraction n is surrounded by air. We must add the extra half wavelength for the front surface reflection, but not for the back surface. The total path difference is then $2d + \frac{1}{2}\lambda_n$. For constructive interference we would have

$$2d + \tfrac{1}{2}\lambda_n = m\lambda_n \qquad m = 1, 2, 3, \ldots \qquad \text{(maxima)}, \tag{41-20}$$

where we have dropped the $m = 0$ solution because it is not physically meaningful. For destructive interference, we obtain

$$2d + \tfrac{1}{2}\lambda_n = (m + \tfrac{1}{2})\lambda_n \qquad m = 0, 1, 2, \ldots \qquad \text{(minima)}. \tag{41-21}$$

These results apply when the index of refraction of the film is greater than the index of refraction of the material on either side—for example, a soap film in air. In this case we include in Eq. 41-19 only the first of the two extra terms. The equations also apply when the index of refraction of the film is less than the index of refraction of the materials on either side—for example, a thin film of air between two glass plates. In this case we include only the second of the two extra terms in Eq. 41-19, as in Sample Problem 41-5. Sample Problem 41-4 illustrates a case in which Eqs. 41-20 and 41-21 *do not* apply.

If the film thickness is not uniform, as in Fig. 41-12 where the film is wedge-shaped, constructive interference occurs in certain parts of the film and destructive interference occurs in others. Bands of maximum and minimum intensity appear, called *fringes of constant thickness.* The width and spacing of the fringes depend on the variation of the film thickness d. If the film is illuminated with white light rather than monochromatic light, the light reflected from various parts of the film is modified by the various constructive or destructive interferences that occur. This accounts for the brilliant colors of soap bubbles and oil slicks.

---

*These statements, which can be proved rigorously from Maxwell's equations, must be modified for light falling on a less dense medium at an angle such that total internal reflection occurs. They must also be modified for reflection from metallic surfaces.

Only if the film is "thin" (d being no more than a few wavelengths of light) is it possible to obtain these types of fringes—that is, fringes that appear localized on the film and are associated with a variable film thickness. For very thick films (say $d \approx 1$ cm), the path difference between the two rays of Fig. 41-13 is many wavelengths, and the phase difference at a given point on the film changes rapidly as we move even a small distance away from a. For "thin" films, however, the phase difference at a also holds for reasonably nearby points; there is a characteristic "patch brightness" for any point on the film, as Fig. 41-12 shows. Interference fringes can be produced for thick films; they are not localized on the film but are at infinity (see Section 41-6).

SAMPLE PROBLEM 41-3. A soap film ($n = 1.33$) in air is 320 nm thick. If it is illuminated with white light at normal incidence, what color will it appear to be in reflected light?

Solution In this problem we want to find the wavelengths λ for which constructive interference occurs. With $\lambda_n = \lambda/n$, we solve Eq. 41-20 for λ and obtain

$$\lambda = \frac{2dn}{m - \frac{1}{2}} = \frac{(2)(320\text{ nm})(1.33)}{m - \frac{1}{2}} = \frac{851\text{ nm}}{m - \frac{1}{2}}.$$

Constructive interference maxima occur for the following wavelengths: 1702 nm ($m = 1$), 567 nm ($m = 2$), 340 nm ($m = 3$), and so forth.

Only the maximum corresponding to $m = 2$ lies in the visible region (between about 400 and 700 nm); light of wavelength 567 nm appears yellow-green. If white light is used to illuminate the film, the yellow-green component is enhanced when viewed by reflection. What is the color of the light transmitted through the film?

SAMPLE PROBLEM 41-4. Lenses are often coated with thin films of transparent substances such as MgF_2 ($n = 1.38$) to reduce the reflection from the glass surface, as in Fig. 41-15. How thick a coating is needed to produce a minimum reflection at the center of the visible spectrum ($\lambda = 550$ nm)?

Solution We assume that the light strikes the lens at near-normal incidence (θ is exaggerated for clarity in Fig. 41-15), and we seek the conditions for destructive interference between rays r_1 and r_2. Equation 41-21 does not apply because in this case a phase change of 180° is associated with *each* ray, for at *both* the front and back surfaces of the MgF_2 film the reflection is from a medium of greater index of refraction. The path difference from Eq. 41-19 is therefore $2d + \frac{1}{2}\lambda_n + \frac{1}{2}\lambda_n$, and the condition for destructive interference is path difference $= (m + \frac{1}{2})\lambda_n$, or

$$2d + \lambda_n = (m + \tfrac{1}{2})\lambda_n.$$

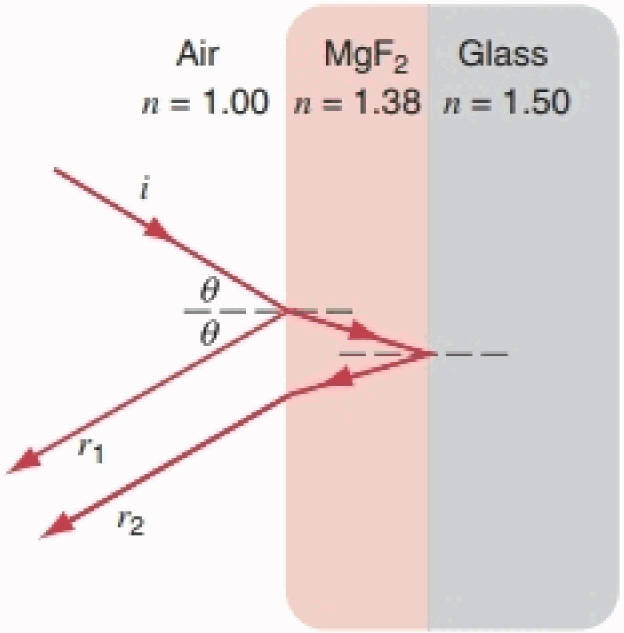


FIGURE 41-15. Sample Problem 41-4. Unwanted reflections from glass can be suppressed (at a chosen wavelength) by coating the glass with a film of the proper thickness.

Putting $\lambda_n = \lambda/n$ and solving for d, we obtain

$$d = \frac{(m - \frac{1}{2})\lambda}{2n} \qquad m = 1, 2, 3, \ldots .$$

We seek the minimum thickness for destructive interference. For $m = 1$, we obtain

$$d = \frac{\lambda}{4n} = \frac{550\text{ nm}}{(4)(1.38)} = 100\text{ nm}.$$

SAMPLE PROBLEM 41-5. Figure 41-16 shows a plano-convex lens of radius of curvature R resting on an accurately plane glass plate and illuminated from above by light of wavelength λ. Figure 41-17 shows that circular interference fringes (called *Newton's rings*) appear, associated with the variable thickness air film between the lens and the plate. Find the radii of the circular interference maxima.

Solution Here it is the ray from the back surface of the (air) film rather than from the front surface that undergoes a phase change of 180°, for it is the one reflected from a medium of higher refractive index. The condition for constructive interference remains

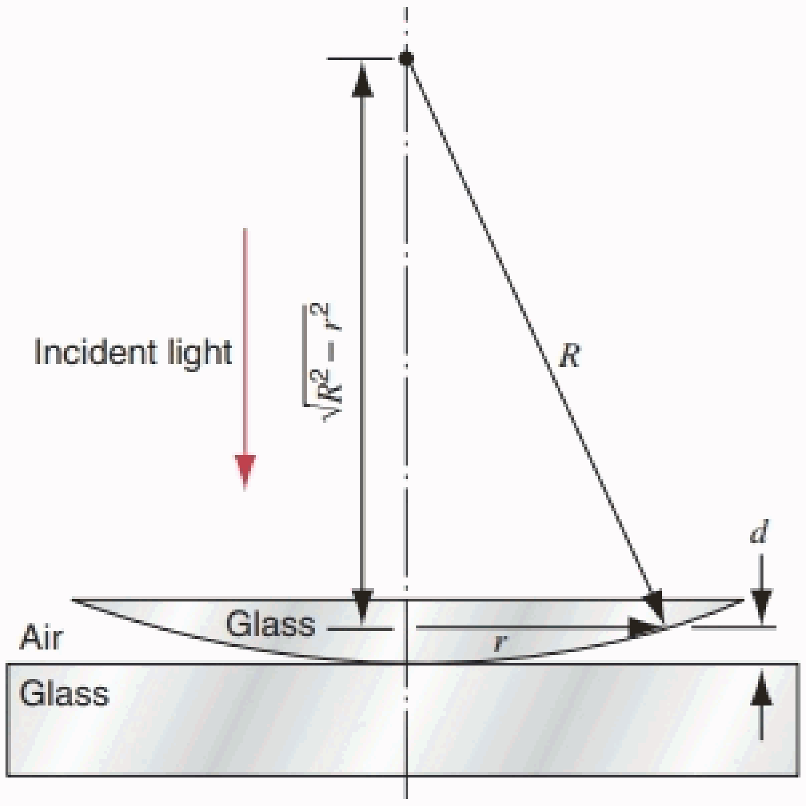


FIGURE 41-16. Sample Problem 41-5. The apparatus for observing Newton's rings.

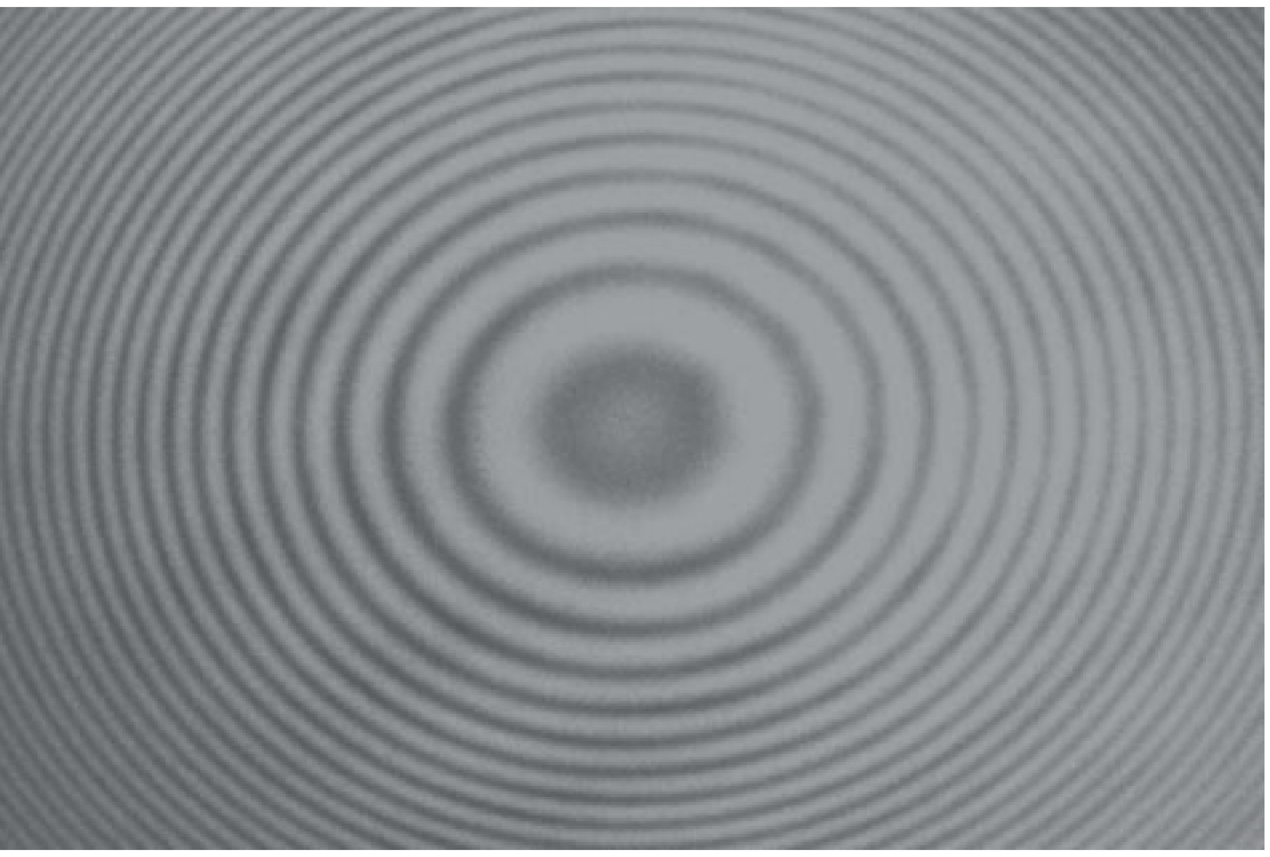


FIGURE 41-17. Circular interference fringes (Newton's rings) observed with the apparatus of Fig. 41-16.

unchanged (Eq. 41-20), however, and is

$$2d = (m - \tfrac{1}{2})\lambda,$$

assuming $n = 1$ for the air film. From Fig. 41-16 we can write

$$d = R - \sqrt{R^2 - r^2} = R - R\left[1 - \left(\frac{r}{R}\right)^2\right]^{1/2}.$$

If $r/R \ll 1$, we can expand the square bracket by the binomial theorem, keeping only two terms, or

$$d = R - R\left[1 - \frac{1}{2}\left(\frac{r}{R}\right)^2 + \cdots\right] \approx \frac{r^2}{2R}.$$

Substituting for d from the interference condition and solving for r, we obtain

$$r = \sqrt{(m - \tfrac{1}{2})\lambda R} \qquad m = 1, 2, \ldots \qquad \text{(maxima)},$$

which gives the radii of the bright rings. If white light is used, each spectrum component produces its own set of circular fringes, and the sets all overlap.

Note that $r > 0$ for $m = 1$. That is, the first bright ring is at $r > 0$, and consequently the center must be dark, as shown in Fig. 41-17. This observation can be taken as further experimental evidence for the 180° phase change on reflection used to obtain Eq. 41-20.

Optical Reversibility and Phase Changes on Reflection (Optional)

G. G. Stokes (1819–1903) used the principle of optical reversibility to investigate the reflection of light at an interface between two media. The principle states that if there is no absorption of light, a light ray that is reflected or refracted will retrace its original path if its direction is reversed. This reminds us that any mechanical system can run backward as well as forward, provided there is no dissipation of energy such as by friction.

Figure 41-18*a* shows a wave of amplitude E reflected and refracted at a surface separating media 1 and 2, where $n_2 > n_1$. The amplitude of the reflected wave is $r_{12}E$, in which r_{12} is an *amplitude reflection coefficient.* The amplitude of the refracted wave is $t_{12}E$, where t_{12} is an *amplitude transmission coefficient.*

The sign of the coefficient indicates the relative phase of the reflected or transmitted component. If we consider only the possibility of phase changes of 0 or 180°, then if $r_{12} = +0.5$, for example, we have a reduction in amplitude on reflection by one-half and no change in phase. For $r_{12} = -0.5$ we have a phase change of 180° because

$$E \sin(\omega t + 180°) = -E \sin \omega t.$$

In Fig. 41-18*b*, the rays indicated by $r_{12}E$ and $t_{12}E$ have been reversed in direction. Ray $r_{12}E$, identified by the single arrows in the figure, is reflected and refracted, producing the rays of amplitudes r_{12}^2E and $r_{12}t_{12}E$. Ray $t_{12}E$, identified by the triple arrows, is also reflected and refracted, producing the rays of amplitudes $t_{12}t_{21}E$ and $t_{12}r_{21}E$ as shown. Note that r_{12} describes a ray in medium 1 reflected from medium 2, and r_{21} describes a ray in medium 2 reflected from medium 1. Similarly, t_{12} describes a ray that passes from medium 1 to medium 2; t_{21} describes a ray that passes from medium 2 to medium 1.

Based on the reversibility principle, we conclude that the two rays in the upper left of Fig. 41-18*b* must be equivalent to the incident ray of Fig. 41-18*a*, reversed; the two rays in the lower left of Fig. 41-18*b* must cancel. This second requirement leads to

$$r_{12}t_{12}E + t_{12}r_{21}E = 0,$$

or

$$r_{12} = -r_{21}.$$

This result tells us that if we compare a wave reflected from medium 1 with one reflected from medium 2, they behave differently in that one or the other undergoes a phase change of 180°.

Experiment shows that the ray reflected from the more optically dense medium suffers the change in phase of 180°. This

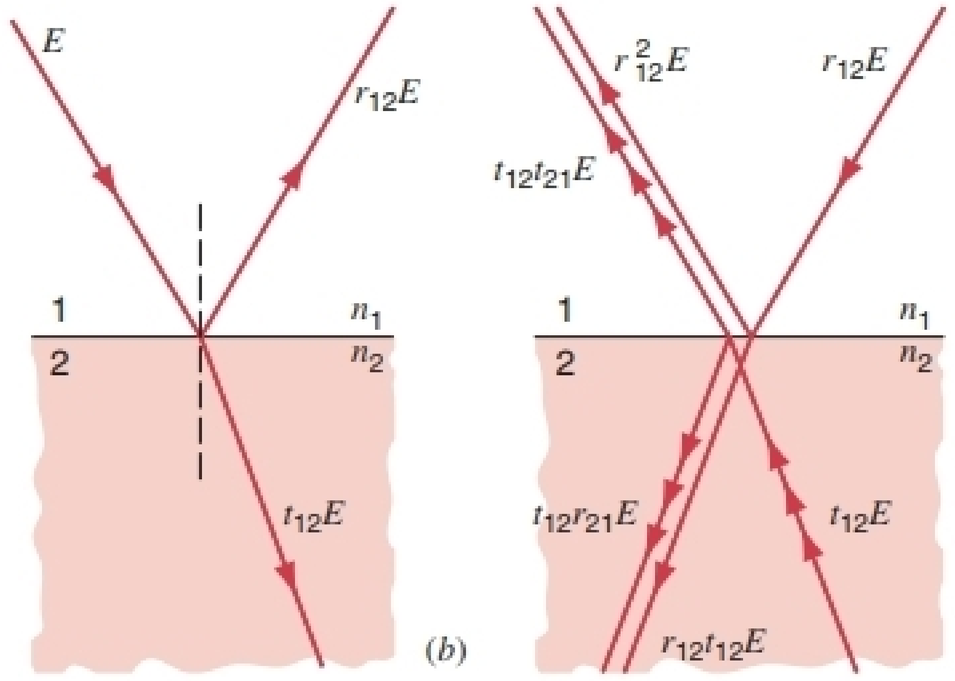


FIGURE 41-18. (*a*) A ray is reflected and refracted at an interface. (*b*) The optically reversed situation; the two rays in the lower left must cancel.

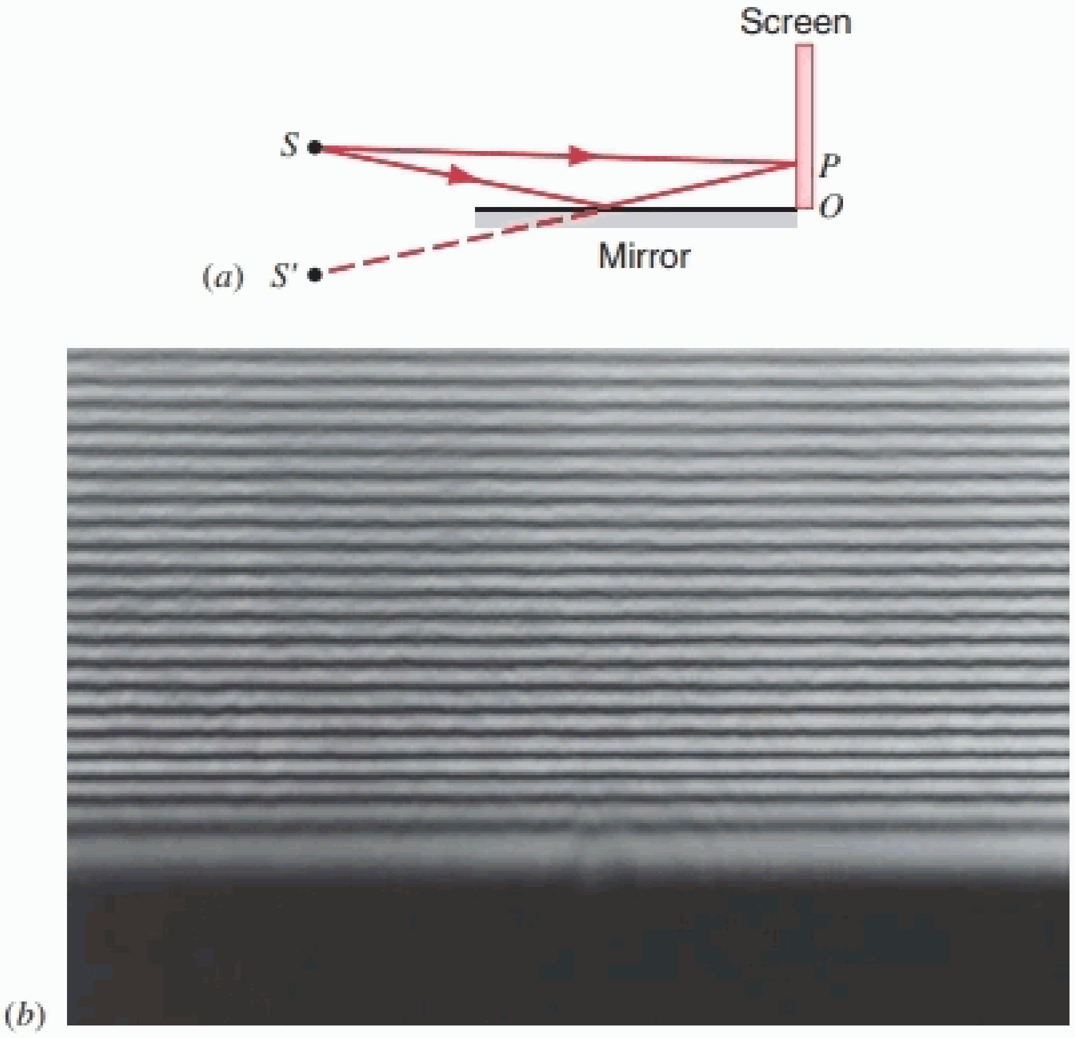


FIGURE 41-19. (*a*) The experimental setup for Lloyd's mirror. Fringes appear on the screen as a result of interference between the direct and reflected beams. (*b*) Fringes observed in the Lloyd's mirror experiment.

can be demonstrated using the setup shown in Fig. 41-19, which is called the Lloyd's mirror experiment. Interference occurs on the screen at an arbitrary point P as a result of the overlap of the direct and reflected beams. We can analyze this experiment as two-source interference, in which one of the sources (S') is the *virtual image* of S in the plane mirror. However, there is one important difference between the apparatus of Fig. 41-19 and the double-slit experiment: the light from the virtual source S' has been reflected from the mirror and has undergone a phase change of 180°. As a result of this phase change, the lower edge of the screen (at O) shows a dark fringe, instead of the bright fringe that appears at the corresponding point (the center of the screen) in the double-slit experiment. Put another way, the appearance of the dark fringe at O shows that one of the interfering beams has been shifted in phase by 180°. Since there is nothing to change the phase of the direct beam SP, it must be the reflected beam that experiences the change in phase. This shows that reflection from a more optically dense medium involves a change in phase of 180°. ■

41-6 MICHELSON'S INTERFEROMETER*

An *interferometer* is a device that can be used to measure lengths or changes in length with great accuracy by means of interference fringes. We describe the form originally built by A. A. Michelson (1852–1931) in 1881.

Consider light that leaves point P on extended source S (Fig. 41-20) and falls on half-silvered mirror M (sometimes called a *beam splitter*). This mirror has a silver coating just thick enough to transmit half the incident light and to reflect half; in the figure we have assumed for convenience that this mirror has negligible thickness. At M the light divides into two waves. One proceeds by transmission a distance d_1 toward mirror M_1; the other proceeds by reflection a distance d_2 toward M_2. The waves are reflected at each of these mirrors and are sent back along their directions of incidence, each wave eventually entering the eye E. Because the waves are coherent, being derived from the same point on the source, they interfere. The eye sees an interference pattern of circular fringes, like those of Newton's rings (Fig. 41-17).

The interference occurs because the two light beams travel different paths between M and M_1 or M_2. Each beam travels its respective path twice (to and from the mirror), so when the beams recombine their path difference is $2(d_2 - d_1)$.

The path difference can be changed by moving mirror M_2, which is mounted on a precise micrometer adjustment. As M_2 is moved, the circular fringes appear to grow or shrink (depending on the direction of motion of M_2), with new rings appearing at the center of the interference pattern and growing outward or larger rings collapsing and disappearing at the center as they shrink. For the center of the fringe pattern to change from bright to dark to bright again, the path difference between the two beams must change by one wavelength, which means (because the light travels the path between M and M_2 twice) that mirror M_2 moves through a distance of $\frac{1}{2}\lambda$.

The interferometer is used to measure changes in length by counting the number of interference fringes that pass the field of view as mirror M_2 is moved. Length measurements made in this way can be accurate if large numbers of fringes are counted.

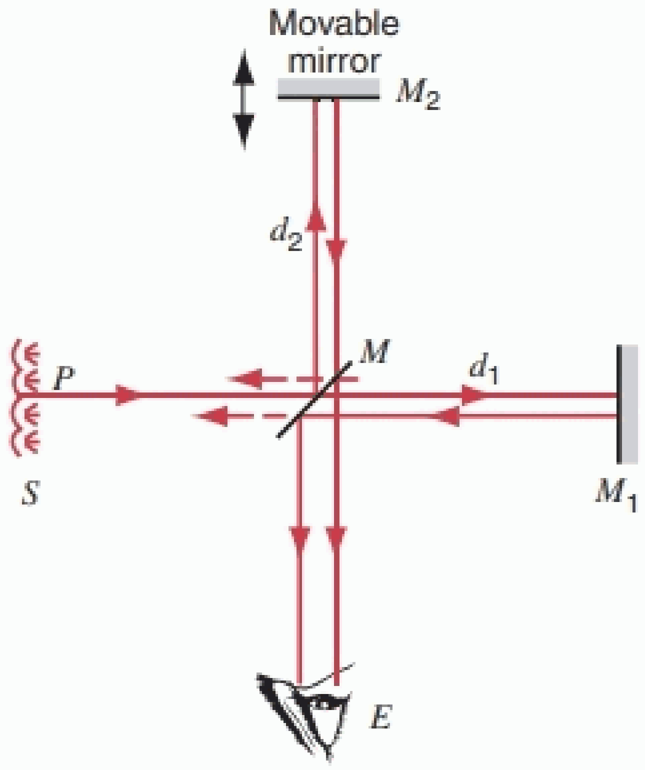


FIGURE 41-20. Michelson's interferometer, showing the path of a ray originating at point P of an extended source S. The ray from P splits at M; the two rays are reflected from mirrors M_1 and M_2 and then recombine at M. Mirror M_2 can be moved to change the path difference between the combining rays.

---

*See "Michelson: America's First Nobel Prize Winner in Science," by R. S. Shankland, *The Physics Teacher,* January 1977, p. 19. See also "Michelson and His Interferometer," by R. S. Shankland, *Physics Today,* April 1974, p. 36.

Michelson measured the length of the standard meter, kept in Paris, in terms of the wavelength of a certain monochromatic red light emitted from a light source containing cadmium. He showed that the standard meter was equivalent to 1,553,163.5 wavelengths of the red cadmium light. For this work he received the Nobel prize in 1907. Michelson's work laid the foundation for the eventual abandonment (in 1961) of the meter bar as a standard of length and for the redefinition of the meter in terms of the wavelength of light. In 1983, as we have seen, even this wavelength standard was not precise enough to meet the growing requirements of science and technology and was replaced by a new standard based on a defined value for the speed of light.

SAMPLE PROBLEM 41-6. Yellow light (λ = 589.00 nm) illuminates a Michelson interferometer. How many bright fringes will be counted as the mirror is moved through 1.0000 cm?

Solution Each fringe corresponds to a movement of the mirror through one-half wavelength. The number of fringes is thus the same as the number of half wavelengths in 1.0000 cm, or

$$\frac{1.0000 \times 10^{-2}\ \text{m}}{\frac{1}{2}(589.00 \times 10^{-9}\ \text{m})} = 33{,}956 \text{ fringes.}$$

Michelson's Interferometer and Light Propagation (Optional)

In Chapter 20, we presented Einstein's hypothesis, now well verified, that in free space light travels with the same speed c no matter what the relative velocity of the source and the observer may be. This hypothesis contradicted the views of 19th-century physicists regarding wave propagation. It was difficult for these physicists, trained as they were in the classical physics of the time, to believe that a wave could be propagated without a medium. If such a medium could be established, the speed c of light would naturally be regarded as the speed *with respect to that medium,* just as the speed of sound always refers to a medium such as air. Although no medium for light propagation was obvious, physicists postulated one, called the *ether,* and hypothesized that its properties were such that it was undetectable by ordinary means such as weighing.

In 1881 (24 years before Einstein's hypothesis), Michelson set himself the task of direct physical verification of the existence of the ether. In particular, Michelson, later joined by E. W. Morley, tried to measure the speed u with which the Earth moves through the ether. Michelson's interferometer was their instrument of choice for this now-famous Michelson–Morley experiment.

The Earth together with the interferometer moving with velocity $\vec{\mathbf{u}}$ through the ether is equivalent to the interferometer at rest with the ether streaming through it with velocity $-\vec{\mathbf{u}}$, as shown in Fig. 41-21. Consider a wave moving along the path MM_1M and one moving along MM_2M. The first corresponds classically to a person rowing a boat a distance d downstream and the same distance upstream; the second corresponds to rowing a boat a distance d across a stream and back.

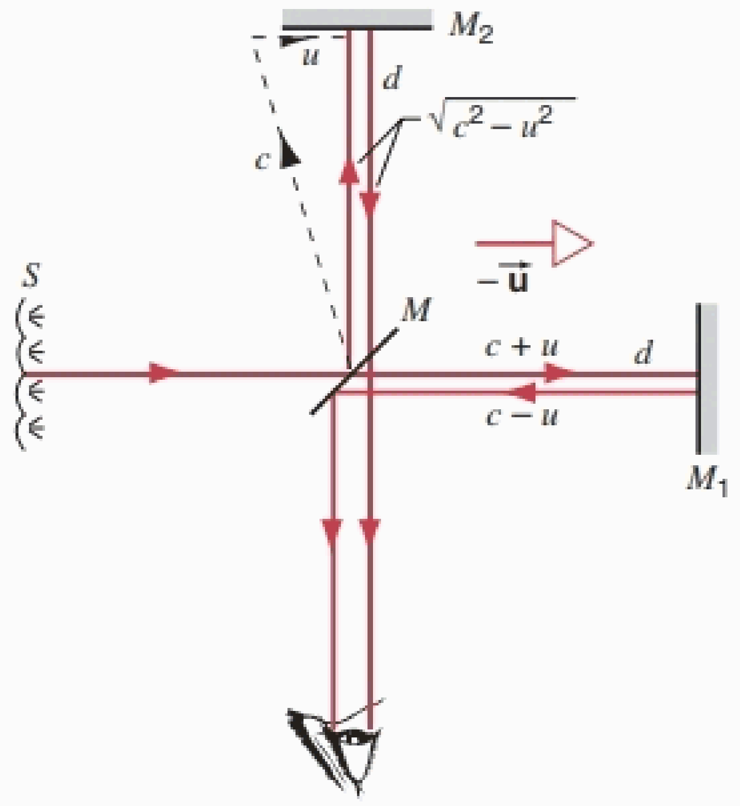


FIGURE 41-21. The "ether" streaming with velocity $-\vec{\mathbf{u}}$ through Michelson's interferometer. The speeds shown are based on the (incorrect) ether hypothesis.

Based on the ether hypothesis the speed of light on the path MM_1 is $c + u$; on the return path M_1M it is $c - u$. The time required for the complete trip is

$$t_1 = \frac{d}{c+u} + \frac{d}{c-u} = d\,\frac{2c}{c^2 - u^2} = \frac{2d}{c}\,\frac{1}{1 - (u/c)^2}.$$

The speed of light, again *based on the ether hypothesis,* for path MM_2 is $\sqrt{c^2 - u^2}$, as Fig. 41-21 suggests. This same speed holds for the return path M_2M, so that the time required for this complete path is

$$t_2 = \frac{2d}{\sqrt{c^2 - u^2}} = \frac{2d}{c}\,\frac{1}{\sqrt{1 - (u/c)^2}}.$$

The difference of time for the two paths is

$$\begin{aligned}\Delta t &= t_1 - t_2\\ &= \frac{2d}{c}\left\{\left[1 - \left(\frac{u}{c}\right)^2\right]^{-1} - \left[1 - \left(\frac{u}{c}\right)^2\right]^{-1/2}\right\}.\end{aligned}$$

Assuming $u/c \ll 1$, we can expand the quantities in the square brackets by using the binomial theorem, retaining only the first two terms. This leads to

$$\begin{aligned}\Delta t &= \frac{2d}{c}\left\{\left[1 + \left(\frac{u}{c}\right)^2 + \cdots\right]\right.\\ &\qquad \left. - \left[1 + \frac{1}{2}\left(\frac{u}{c}\right)^2 + \cdots\right]\right\}\\ &= \frac{2d}{c}\left\{\frac{1}{2}\left(\frac{u}{c}\right)^2\right\} = \frac{du^2}{c^3}. \qquad (41\text{-}22)\end{aligned}$$

Now let the entire interferometer be rotated through 90°. This interchanges the roles of the two light paths,

MM_1M now being the "cross-stream" path and MM_2M the "downstream and upstream" path. The time difference between the two waves entering the eye is also reversed; this changes the phase difference between the combining waves and alters the positions of the interference maxima. The experiment consists of looking for a shift of the interference fringes as the apparatus is rotated.

The *change* in time difference is $2\Delta t$, which corresponds to a phase difference $\Delta\phi = \omega(2\Delta t)$, where $\omega\ (=2\pi c/\lambda)$ is the angular frequency of the light wave. The expected maximum shift in the number of fringes on a 90° rotation is

$$\Delta N = \frac{\Delta\phi}{2\pi} = \frac{\omega(2\Delta t)}{2\pi} = \frac{2c\Delta t}{\lambda} = \frac{2d}{\lambda}\left(\frac{u}{c}\right)^2, \quad (41\text{-}23)$$

where we have used Eq. 41-22 for Δt.

In the Michelson–Morley interferometer let $d = 11$ m (obtained by multiple reflection in the interferometer) and $\lambda = 5.9 \times 10^{-7}$ m. If u is assumed to be roughly the orbital speed of the Earth, then $u/c \approx 10^{-4}$. The expected maximum fringe shift when the interferometer is rotated through 90° is then

$$\Delta N = \frac{2d}{\lambda}\left(\frac{u}{c}\right)^2 = \frac{(2)(11\ \text{m})}{5.9 \times 10^{-7}\ \text{m}}(10^{-4})^2 = 0.4.$$

Even though a shift of only about 0.4 of a fringe was expected, Michelson and Morley were confident that they could observe a shift of 0.01 fringe. *They found from their experiment, however, that there was no observable fringe shift!*

The analogy between a light wave in the supposed ether and a boat moving in water, which seemed so evident in 1881, is simply incorrect. The derivation based on this analogy is incorrect for light waves. When the analysis is carried through based on Einstein's hypothesis, the observed null result is clearly predicted, the speed of light being c for all paths. The motion of the Earth around the Sun and the rotation of the interferometer have, in Einstein's view, no effect whatever on the speed of the light waves in the interferometer.

It should be made clear that although Einstein's hypothesis is completely consistent with the null result of the Michelson–Morley experiment, this experiment standing alone does not serve as a proof for Einstein's hypothesis. Einstein said that no number of experiments, however large, could prove him right but that a single experiment could prove him wrong. Our present-day belief in Einstein's hypothesis rests on consistent agreement in a large number of experiments designed to test it. The "single experiment" that might prove Einstein wrong has never been found.

MULTIPLE CHOICE

41-1 Two-Source Interference

41-2 Double-Slit Interference

1. What is the minimum slit separation d for interference to produce at least one maximum other than the central maximum?
(A) $d \geq 5\lambda$ (B) $d \geq 3\lambda$ (C) $d \geq \lambda$ (D) $d \geq \lambda/3$
(E) There is no minimum.

2. What is the maximum slit separation d for which any interference fringes will be produced?
(A) There is no maximum.
(B) $d \leq 5\lambda$ (C) $d \leq 3\lambda$ (D) $d \leq \lambda$ (E) $d \leq \lambda/3$

3. A double-slit apparatus is set up as shown in Fig. 41-5. Green light is used to form the interference pattern. The green light is then replaced by red light. Which of the following changes will allow the spacing of the bright bands on the screen to remain unaffected by the change of light color?
(A) Increase D. (B) Decrease d.
(C) Increase d and/or decrease D.
(D) Increase the brightness of the light source.

4. Suppose a filter is placed over one of the slits in Fig. 41-3 to prevent light from passing through.

(*a*) What would happen to the intensity of the light at the location of central maximum in Fig. 41-4?
(A) It increases. (B) It decreases.
(C) It remains the same.
(D) It depends on which slit was covered.

(*b*) What would happen to the intensity of the light at the location of the first minimum?
(A) It increases. (B) It decreases.
(C) It remains the same.
(D) It depends on which slit was covered.

(*c*) What would happen to the intensity of the light at the location of the first maximum next to the central maximum?
(A) It increases. (B) It decreases.
(C) It remains the same.
(D) It depends on which slit was covered.

5. A double-slit apparatus is used to observe an interference pattern projected onto a screen from a stationary light source. If the light source is instead moved toward the double slits at constant speed, what will be observed on the screen?
(A) The interference pattern will remain unchanged.
(B) The fringes will move farther apart to a new fixed distance.
(C) The fringes will move closer together to a new fixed distance.
(D) The fringes will continue to move farther and farther apart as the source is brought closer.

6. A physics instructor slips a thin sheet of a transparent material with an index of refraction only slightly larger than air over one of the slits in a double-slit demonstration. What will happen to the interference pattern on the screen?
(A) There will be no change.
(B) The fringes will spread further apart.
(C) The fringes will move closer together.
(D) The fringes will shift position, but not change spacing.

7. A physics instructor slips a thin sheet of a transparent material with an index of refraction only slightly larger than air over *both* of the slits in a double-slit demonstration. What will be observed with the interference pattern on the screen?
- (A) There will be no change.
- (B) The fringes will spread further apart.
- (C) The fringes will move closer together.
- (D) The fringes will shift position, but not change spacing.

8. Two closely located, monochromatic light sources create a fixed interference pattern on a screen. The frequency of one of the sources is then decreased. What will happen to the interference pattern as a result?
- (A) The pattern will completely disappear.
- (B) The fringe spacing will increase.
- (C) The fringes will shift left or right, but the spacing will remain constant.
- (D) The fringe spacing will decrease.

9. Equation 41-4 is an approximation for small angles; in this approximation the spacing between adjacent maxima is a constant given by Δy. In reality, the spacing between adjacent maxima
- (A) is smaller than Δy, getting even smaller as θ increases.
- (B) is smaller than Δy, but gets larger as θ increases.
- (C) is larger than Δy, but gets smaller as θ increases.
- (D) is larger than Δy, getting even larger as θ increases.

41-3 Coherence

10. Which of the following must be the same for two rays of light to be considered coherent? (Choose all that apply.)

(A) Frequency (B) Wavelength (C) Wave speed
(D) Wave phase (E) Wave amplitude

41-4 Intensity in Double-Slit Interference

11. For fun, a student constructs a *three*-slit interference experiment. If the light from each of the three slits arrives in phase at the central maximum, then how will the intensity I compare to the intensity I_0 from a single slit?

(A) $I = I_0$ (B) $I = 3I_0$ (C) $I = 6I_0$ (D) $I = 9I_0$

12. Three coherent, equal intensity light rays arrive at a point P on a screen to produce an interference minimum of zero intensity. If any two of the rays are blocked, the intensity of the light at P is I_1. What is the intensity of the light at P if only one of the rays is blocked?

(A) 0 (B) $I_1/2$ (C) I_1 (D) $2I_1$ (E) $4I_1$

41-5 Interference from Thin Films

13. A ray of light is incident on a thin film. Two of the reflected rays are shown, and two of the transmitted rays are shown in Fig. 41-22.

(*a*) Assume that rays a and b undergo a phase change because of differences in the indices of refraction. Which would be the proper ordering of the indices?

(A) $n_1 > n_2 > n_3$ (B) $n_2 > n_3 > n_1$
(C) $n_3 > n_2 > n_1$ (D) $n_3 > n_1 > n_2$
(E) None of these; the specified phase changes cannot occur.

(*b*) Assume that rays a and c undergo a phase change because of differences in the indices of refraction. Which would be the proper ordering of the indices?

(A) $n_1 > n_2 > n_3$ (B) $n_2 > n_3 > n_1$
(C) $n_3 > n_2 > n_1$ (D) $n_3 > n_1 > n_2$
(E) None of these; the specified phase changes cannot occur.

(*c*) Assume that rays a and d undergo a phase change because of differences in the indices of refraction. Which would be the proper ordering of the indices?

(A) $n_1 > n_2 > n_3$ (B) $n_2 > n_3 > n_1$
(C) $n_3 > n_2 > n_1$ (D) $n_3 > n_1 > n_2$
(E) None of these; the specified phase changes cannot occur.

(*d*) Assume that rays b and c undergo a phase change because of differences in the indices of refraction. Which would be the proper ordering of the indices?

(A) $n_1 > n_2 > n_3$ (B) $n_2 > n_3 > n_1$
(C) $n_3 > n_2 > n_1$ (D) $n_3 > n_1 > n_2$
(E) None of these; the specified phase changes cannot occur.

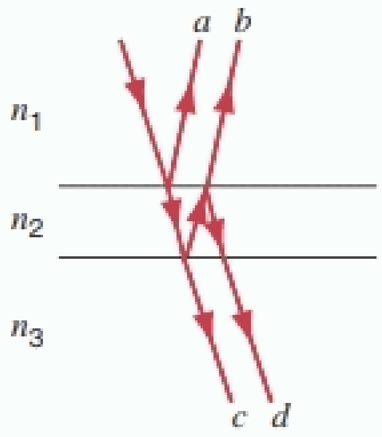


FIGURE 41-22. Multiple-choice question 13.

14. Why do we not see the effects of thin-film interference when looking at the pane of glass in a window?
- (A) Interference does not happen in thick films.
- (B) The fringes are too dim.
- (C) The fringes are too far apart.
- (D) The bright and dark fringes tend to overlap.

15. 550-nm light strikes a thin film normal to the surface; all of the light is transmitted and none is reflected. A second ray of light of wavelength λ strikes the same thin film at a very small angle to the normal; all of the light is also transmitted and none is reflected. What can be concluded about λ?

(A) $\lambda > 550$ nm (B) $\lambda = 550$ nm (C) $\lambda < 550$ nm
(D) Nothing without knowledge of the indices of refraction of the various substances in the problem

41-6 Michelson's Interferometer

QUESTIONS

1. Is Young's experiment an interference experiment or a diffraction experiment, or both?

2. In Young's double-slit interference experiment, using a monochromatic laboratory light source, why is screen *A* in Fig. 41-7 necessary? If the source of light is a laser beam, screen *A* is not needed. Why?

3. What changes, if any, occur in the pattern of interference fringes if the apparatus of Fig. 41-5 is placed under water?

4. Do interference effects occur for sound waves? Recall that sound is a longitudinal wave and that light is a transverse wave.

5. It is not possible to show interference effects between light from two separate sodium vapor lamps but you can show interference effects between sound from two loudspeakers that are driven by separate oscillators. Explain why this is so.

6. If interference between light waves of different frequencies were possible, one should observe light beats, just as one obtains sound beats from two sources of sound with slightly different frequencies. Discuss how one might experimentally look for this possibility.

7. Why are parallel slits preferable to the pinholes that Young used in demonstrating interference?

8. Is coherence important in reflection and refraction?

9. Describe the pattern of light intensity on screen *C* in Fig. 41-5 if one slit is covered with a red filter and the other with a blue filter, with the incident light being white.

10. If one slit in Fig. 41-5 is covered, what change would occur in the intensity of light at the center of the screen?

11. We are all bathed continuously in electromagnetic radiation, from the Sun, from radio and TV signals, from the stars and other celestial objects. Why do these waves not interfere with each other?

12. In calculating the disturbance produced by a pair of superimposed wavetrains, when should you add intensities and when amplitudes?

13. In Young's double-slit experiment, suppose that screen *A* in Fig. 41-7 contained two very narrow, parallel slits instead of one. (*a*) Show that if the spacing between these slits is properly chosen, the interference fringes can be made to disappear. (*b*) Under these conditions, would you call the beams emerging from slits S_1 and S_2 in screen *B* coherent? They do not produce interference fringes. (*c*) Discuss what would happen to the interference fringes in the case of a single slit in screen *A* if the slit width were gradually increased.

14. Defend this statement: Fig. 41-8*a* is a sine (or cosine) wave but Fig. 41-8*b* is not. Indeed, you cannot assign a unique frequency to the curve of Fig. 41-8*b*. Why not? (Hint: Think of Fourier analysis.)

15. Most of us are familiar with rotating or oscillating radar antennas that produce rotating or oscillating beams of microwave radiation. It is also possible to produce an oscillating beam of microwave radiation without any mechanical motion of the transmitting antenna. This is done by periodically changing the phase of the radiation as it emerges from various sections of the (long) transmitting antenna. Convince yourself that, by constructive interference from various parts of the fixed antenna, an oscillating microwave beam can indeed be so produced.

16. What causes the fluttering of radio reception when an airplane flies overhead?

17. Is it possible to have coherence between light sources that are emitting light of different wavelengths?

18. An automobile directs its headlights onto the side of a barn. Why are interference fringes not produced in the region in which light from the two beams overlaps?

19. Suppose that the film coating in Fig. 41-15 had an index of refraction greater than that of the glass. Could it still be nonreflecting? If so, what difference would the coating make?

20. What are the requirements for maximum intensity when viewing a thin film by transmitted light?

21. Why does a film (for example, a soap bubble or an oil slick) have to be "thin" to display interference effects? Or does it? How thin is "thin"?

22. Why do coated lenses (see Sample Problem 41-4) look purple by reflected light?

23. Ordinary windows reflect light from both their interior and exterior plane surfaces. Why then do we not see interference effects?

24. If you wet your eyeglasses to clean them, you will notice that as the water evaporates the glasses become markedly less reflecting for a short time. Explain why.

25. A lens is coated to reduce reflection, as in Sample Problem 41-4. What happens to the energy that had previously been reflected? Is it absorbed by the coating?

26. Consider the following objects that produce colors when exposed to sunlight: (1) soap bubbles, (2) rose petals, (3) the inner surface of an oyster shell, (4) thin oil slicks, (5) nonreflecting coatings on camera lenses, and (6) peacock tail feathers. The colors displayed by all but one of these are purely interference phenomena, with no pigments involved. Which one is the exception? Why do the others seem to be "colored"?

27. A soap film on a wire loop held in air appears black at its thinnest portion when viewed by reflected light. On the other hand, a thin oil film floating on water appears bright at its thinnest portion when similarly viewed from the air above. Explain these phenomena.

28. Very small changes in the angle of incidence do not change the interference conditions much for "thin" films but they do change them for "thick" films. Why?

29. An optical flat is a slab of glass that has been ground flat to within a small fraction of a wavelength. How may it be used to test the flatness of a second slab of glass?

30. In a Newton's rings experiment, is the central spot, as seen by reflection, dark or light? Explain.

31. In connection with the phase change on reflection at an interface between two transparent media, do you think that phase shifts other than 0 or π are possible? Do you think that phase shifts can be calculated rigorously from Maxwell's equations?

32. The directional characteristics of a certain radar antenna as a receiver of radiation are known. What can be said about its directional characteristics as a transmitter?

33. A person in a dark room, looking through a small window, can see a second person standing outside in bright sunlight. The second person cannot see the first person. Is this a failure of the principle of optical reversibility? Assume there is no absorption of light.

34. Why is it necessary to rotate the interferometer in the Michelson–Morley experiment?

35. How is the negative result of the Michelson–Morley experiment interpreted according to Einstein's theory of relativity?

36. If the path length to the movable mirror in Michelson's interferometer (see Fig. 41-20) greatly exceeds that to the fixed mirror (say, by more than a meter) the fringes begin to disappear. Explain why. Lasers greatly extend this range. Why?

37. How would you construct an acoustical Michelson interferometer to measure wavelengths of sound? Discuss differences from the optical interferometer.

EXERCISES

41-1 Two-Source Interference

41-2 Double-Slit Interference

1. Monochromatic green light, wavelength = 554 nm, illuminates two parallel, narrow slits 7.7 μm apart. Calculate the angular position of the third-order ($m = 3$) bright fringe (*a*) in radians and (*b*) in degrees.

2. Monochromatic light illuminates two parallel slits a distance d apart. The first maximum is observed at an angular position of 15°. By what percentage should d be increased or decreased so that the *second* maximum will instead be observed at 15°?

3. A double-slit experiment is performed with blue-green light of wavelength 512 nm. The slits are 1.2 mm apart and the screen is 5.4 m from the slits. How far apart are the bright fringes as seen on the screen?

4. Find the slit separation of a double-slit arrangement that will produce bright interference fringes 1.00° apart in angular separation. Assume a wavelength of 592 nm.

5. A double-slit arrangement produces interference fringes for sodium light ($\lambda = 589$ nm) that are 0.23° apart. For what wavelength would the angular separation be 10% greater? Assume that the angle θ is small.

6. A double-slit arrangement produces interference fringes for sodium light ($\lambda = 589$ nm) that are 0.20° apart. What is the angular fringe separation if the entire arrangement is immersed in water ($n = 1.33$)?

7. In a double-slit experiment, the distance between slits is 5.22 mm and the slits are 1.36 m from the screen. Two interference patterns can be seen on the screen, one due to light with wavelength 480 nm and the other due to light with wavelength 612 nm. Find the separation on the screen between the third-order interference fringes of the two different patterns.

8. In an interference experiment in a large ripple tank (see Fig. 41-2), the coherent vibrating sources are placed 120 mm apart. The distance between maxima 2.0 m away is 180 mm. If the speed of ripples is 25 cm/s, calculate the frequency of the vibrating sources.

9. If the distance between the first and tenth minima of a double-slit pattern is 18 mm and the slits are separated by 0.15 mm with the screen 50 cm from the slits, what is the wavelength of the light used?

10. Two coherent radio point sources separated by 2.0 m are radiating in phase with $\lambda = 0.50$ m. A detector moved in a circular path around the two sources in a plane containing them will show how many maxima?

11. Sketch the interference pattern expected from using two pinholes rather than narrow slits.

12. Sodium light ($\lambda = 589$ nm) falls on a double slit of separation $d = 0.180$ mm. A thin lens ($f = 1.13$ m) is placed near the slit as in Fig. 41-6. What is the linear fringe separation on a screen placed in the focal plane of the lens?

13. Sodium light ($\lambda = 589$ nm) falls on a double slit of separation $d = 2.0$ mm. The slit-screen distance D is 40 mm. What fractional error is made by using Eq. 41-1 to locate the tenth bright fringe on the screen?

41-3 Coherence

14. The coherence length of a wavetrain is the distance over which the phase constant is the same. (*a*) If an individual atom emits coherent light for 1×10^{-8} s, what is the coherence length of the wavetrain? (*b*) Suppose a partially reflecting mirror separates this wave train into two parts that are later reunited after one beam travels 5 m and the other 10 m. Do the waves produce interference fringes observable by a human eye?

41-4 Intensity in Double-Slit Interference

15. Source A of long-range radio waves leads source B by 90°. The distance r_A to a detector is greater than the distance r_B by 100 m. What is the phase difference at the detector? Both sources have a wavelength of 400 m.

16. Find the phase difference between the waves from the two slits arriving at the mth dark fringe in a double-slit experiment.

17. Light of wavelength 600 nm is incident normally on two parallel, narrow slits separated by 0.60 mm. Sketch the intensity pattern observed on a distant screen as a function of angle θ for the range of values $0 \leq \theta \leq 0.0040$ radians.

18. Find the sum of the following quantities (*a*) graphically, using phasors; and (*b*) using trigonometry:

$$y_1 = 10 \sin \omega t,$$

$$y_2 = 8.0 \sin (\omega t + 30°).$$

19. S_1 and S_2 in Fig. 41-23 are effective point sources of radiation, excited by the same oscillator. They are coherent and in phase with each other. Placed a distance $d = 4.17$ m apart, they emit equal amounts of power in the form of 1.06-m-wavelength electromagnetic waves. (*a*) Find the positions of the first, second, and third maxima of the received signal, as the detector D

is moved along the x axis starting at O. (b) Is the intensity at the nearest minimum equal to zero? Justify your answer.

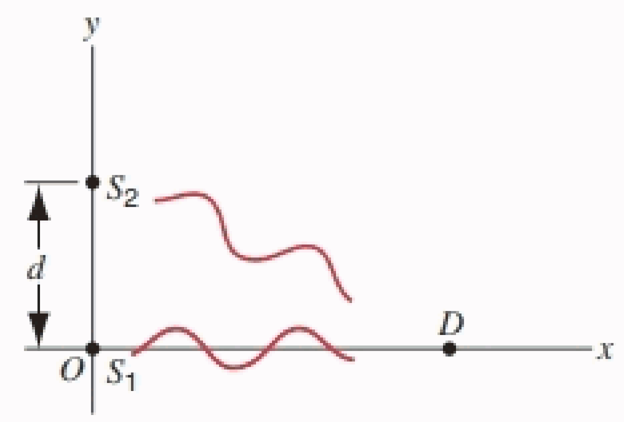


FIGURE 41-23. Exercise 19.

20. Add the following quantities graphically, using the phasor method (see Sample Problem 41-2), and algebraically:

$$y_1 = 10 \sin \omega t,$$

$$y_2 = 14 \sin (\omega t + 26°),$$

$$y_3 = 4.7 \sin (\omega t - 41°).$$

41-5 Interference from Thin Films

21. We wish to coat a flat slab of glass ($n = 1.50$) with a transparent material ($n = 1.25$) so that light of wavelength 620 nm (in vacuum) incident normally is not reflected. What minimum thickness could the coating have?

22. A thin film in air is 410 nm thick and is illuminated by white light normal to its surface. Its index of refraction is 1.50. What wavelengths within the visible spectrum will be intensified in the reflected beam?

23. A disabled tanker leaks kerosene ($n = 1.20$) into the Persian Gulf, creating a large slick on top of the water ($n = 1.33$). (a) If you are looking straight down from an airplane onto a region of the slick where its thickness is 460 nm, for which wavelength(s) of visible light is the reflection the greatest? (b) If you are scuba diving directly under this same region of the slick, for which wavelength(s) of visible light is the transmitted intensity the strongest?

24. In costume jewelry, rhinestones (made of glass with $n = 1.5$) are often coated with silicon monoxide ($n = 2.0$) to make them more reflective. How thick should the coating be to achieve strong reflection for 560-nm light, incident normally?

25. If the wavelength of the incident light is $\lambda = 572$ nm, rays A and B in Fig. 41-24 are out of phase by 1.50λ. Find the thickness d of the film.

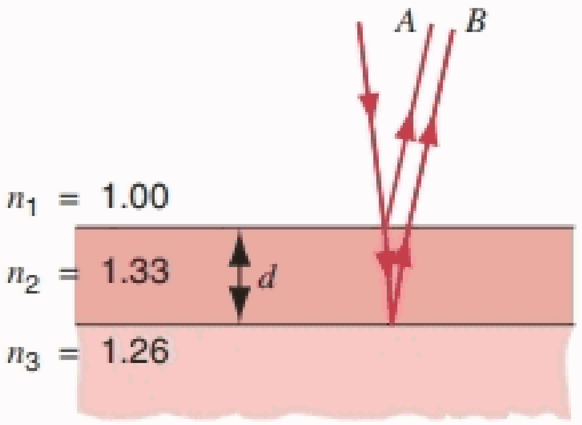


FIGURE 41-24. Exercise 25.

26. Light of wavelength 585 nm is incident normally on a thin, soapy film ($n = 1.33$) suspended in air. If the film is 0.00121 mm thick, determine whether it appears bright or dark when observed from a point near the light source.

27. A thin film of acetone (index of refraction = 1.25) is coating a thick glass plate (index of refraction = 1.50). Plane light waves of variable wavelengths are incident normal to the film. When one views the reflected wave, it is noted that complete destructive interference occurs at 600 nm and constructive interference at 700 nm. Calculate the thickness of the acetone film.

28. White light reflected at perpendicular incidence from a soap film in air has, in the visible spectrum, an interference maximum at 600 nm and a minimum at 450 nm with no minimum in between. If $n = 1.33$ for the film, what is the film thickness, assumed to be uniform?

29. A broad source of light ($\lambda = 680$ nm) illuminates normally two glass plates 120 mm long that touch at one end and are separated by a wire 0.048 mm in diameter at the other end (Fig. 41-25). How many bright fringes appear over the 120-mm distance?

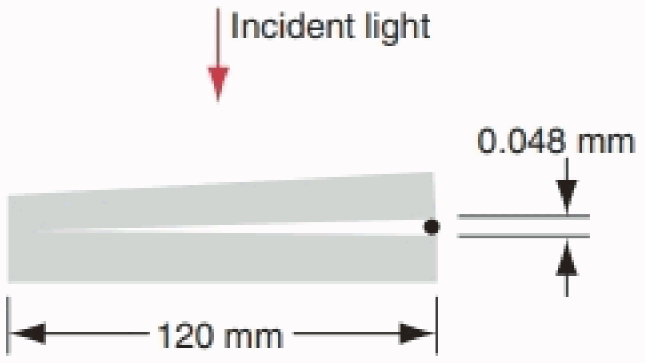


FIGURE 41-25. Exercise 29.

30. An oil drop ($n = 1.20$) floats on a water ($n = 1.33$) surface and is observed from above by reflected light (see Fig. 41-26). (a) Will the outer (thinnest) regions of the drop correspond to a bright or a dark region? (b) How thick is the oil film where one observes the third blue region from the outside of the drop? (c) Why do the colors gradually disappear as the oil thickness becomes larger?

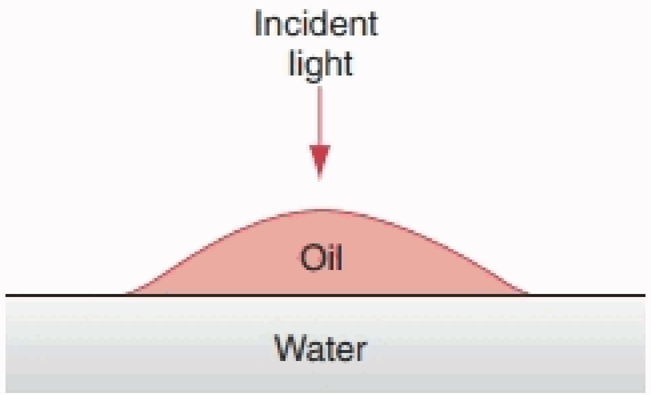


FIGURE 41-26. Exercise 30.

31. Light of wavelength 630 nm is incident normally on a thin, wedge-shaped film with index of refraction 1.50. There are ten bright and nine dark fringes over the length of the film. By how much does the film thickness change over this length?

32. In an air wedge formed by two plane glass plates that are touching each other along one edge, there are 4001 dark lines observed when viewed by reflected monochromatic light. When the air between the plates is evacuated, only 4000 such lines are observed. Calculate the index of refraction of the air from these data.

33. In a Newton's rings experiment, the radius of curvature R of the lens is 5.0 m and its diameter is 20 mm. (*a*) How many rings are produced? (*b*) How many rings would be seen if the arrangement were immersed in water ($n = 1.33$)? Assume that $\lambda = 589$ nm.

34. The diameter of the tenth bright ring in a Newton's rings apparatus changes from 1.42 to 1.27 cm as a liquid is introduced between the lens and the plate. Find the index of refraction of the liquid.

35. A Newton's rings apparatus is used to determine the radius of curvature of a lens. The radii of the nth and $(n + 20)$th bright rings are measured and found to be 0.162 cm and 0.368 cm, respectively, in light of wavelength 546 nm. Calculate the radius of curvature of the lower surface of the lens.

36. A ship approaching harbor is transmitting at a wavelength of $\lambda = 3.43$ m from its antenna located $h = 23$ m above sea level. The receiving station antenna is located $H = 160$ m above sea level. What is the horizontal distance D between ship and receiving tower when radio contact is momentarily lost for the first time? Assume that the calm ocean reflects radio waves perfectly according to the law of reflection. See Fig. 41-27.

37. In Sample Problem 41-4, assume that there is zero reflection for light of wavelength 550 nm at normal incidence. Calculate the factor by which the reflection is diminished by the coating at (*a*) 450 nm and (*b*) 650 nm. (Hint: Calculate ϕ in Eq. 41-16.)

41-6 Michelson's Interferometer

38. A thin film with $n = 1.42$ for light of wavelength 589 nm is placed in one arm of a Michelson interferometer. If a shift of 7.0 fringes occurs, what is the film thickness?

39. If mirror M_2 in Michelson's interferometer is moved through 0.233 mm, 792 fringes are counted with a light meter. What is the wavelength of the light?

40. An airtight chamber 5.0 cm long with glass windows is placed in one arm of a Michelson interferometer as indicated in Fig. 41-28. Light of wavelength $\lambda = 500$ nm is used. The air is slowly evacuated from the chamber using a vacuum pump. While the air is being removed, 60 fringes are observed to pass through the view. From these data, find the index of refraction of air at atmospheric pressure.

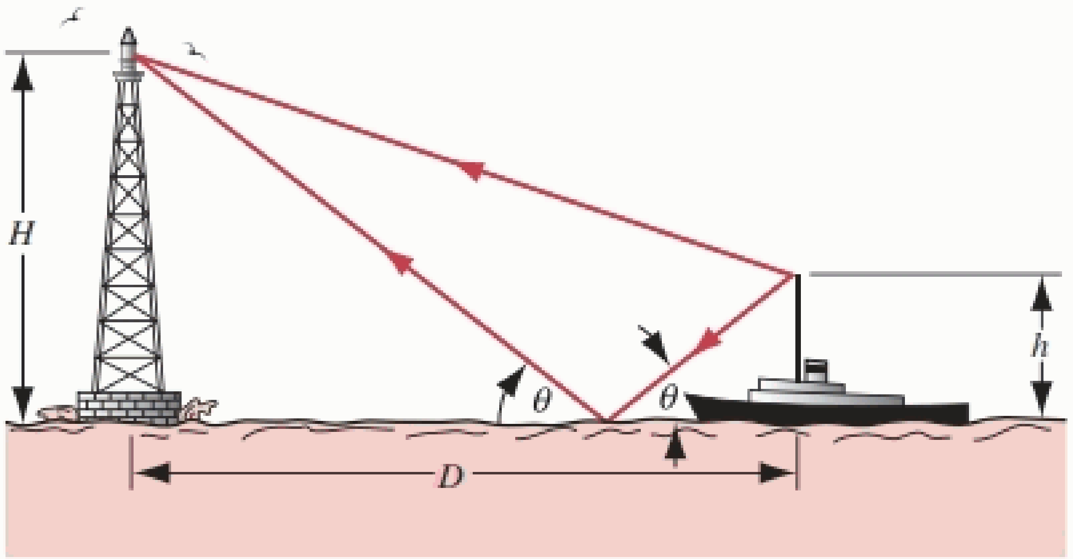


FIGURE 41-27. Exercise 36.

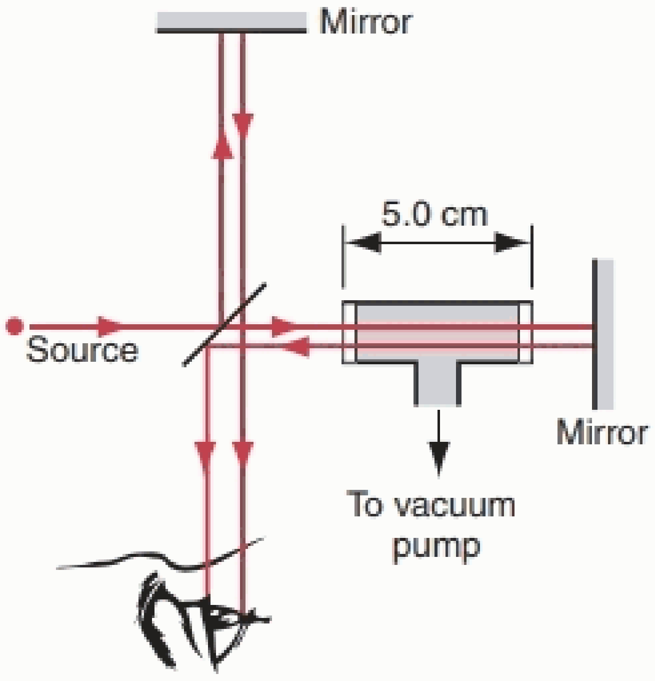


FIGURE 41-28. Exercise 40.

PROBLEMS

1. In the front of a lecture hall, a coherent beam of monochromatic light from a helium–neon laser ($\lambda = 632.8$ nm) illuminates a double slit. From there it travels a distance of 20.0 m to a mirror at the back of the hall, and returns the same distance to a screen. (*a*) In order that the distance between interference maxima be 10.0 cm, what should be the distance between the two slits? (*b*) State what you will see if the lecturer slips a thin sheet of cellophane over one of the slits. The path through the cellophane contains 2.5 more waves than a path through air of the same geometric thickness.

2. A thin flake of mica ($n = 1.58$) is used to cover one slit of a double-slit arrangement. The central point on the screen is occupied by what used to be the seventh bright fringe. If $\lambda = 550$ nm, what is the thickness of the mica?

3. Two point sources, S_1 and S_2 in Fig. 41-29, emit coherent waves. Show that curves such as that given, over which the phase difference for rays r_1 and r_2 is a constant, are hyperbolas. (Hint: A constant phase difference implies a constant difference in length between r_1 and r_2.) The OMEGA system of sea navigation relies on this principle. S_1 and S_2 are phase-locked transmitters. The ship's navigator notes the received phase difference on an oscilloscope and locates the ship on a hyperbola. Reception of signals from a third transmitter is needed to determine the position on that hyperbola.

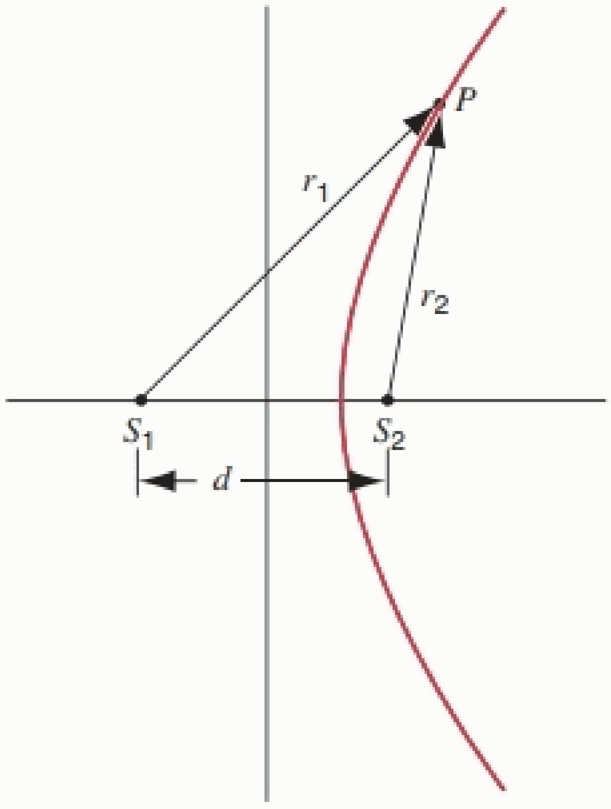


FIGURE 41-29. Problem 3.

4. One slit of a double-slit arrangement is covered by a thin glass plate of index of refraction 1.4, and the other by a thin glass plate of index of refraction 1.7. The point on the screen where the central maximum fell before the glass plates were inserted is now occupied by what had been the $m = 5$ bright fringe before. Assume that $\lambda = 480$ nm and that the plates have the same thickness t, and find the value of t.

5. Show that the half-width $\Delta\theta$ of the double-slit interference fringes is given by

$$\Delta\theta = \frac{\lambda}{2d}$$

if θ is small enough so that $\sin\theta \approx \theta$. The half-width is the angle between the two points in the fringe where the intensity is one-half that at the center of the fringe.

6. One of the slits of a double-slit system is wider than the other, so that the amplitude of the light reaching the central part of the screen from one slit, acting alone, is twice that from the other slit, acting alone. Derive an expression for the intensity I in terms of θ.

7. A plane wave of monochromatic light falls normally on a uniformly thin film of oil that covers a glass plate. The wavelength of the source can be varied continuously. Complete destructive interference of the reflected light is observed for wavelengths of 485 and 679 nm and for no wavelengths between them. If the index of refraction of the oil is 1.32 and that of the glass is 1.50, find the thickness of the oil film.

8. A sheet of glass having an index of refraction of 1.40 is to be coated with a film of material having an index of refraction of 1.55 such that green light (wavelength = 525 nm) is preferentially transmitted. (*a*) What is the minimum thickness of the film that will achieve the result? (*b*) Why are other parts of the visible spectrum not also preferentially transmitted? (*c*) Will the transmission of any colors be sharply reduced?

9. Two pieces of plate glass are held together in such a way that the air space between them forms a very thin wedge. Light of wavelength 480 nm strikes the upper surface perpendicularly and is reflected from the lower surface of the top glass and the upper surface of the bottom glass, thereby producing a series of interference fringes. How much thicker is the air wedge at the sixteenth fringe than it is at the sixth?

10. A perfectly flat piece of glass ($n = 1.50$) is placed over a perfectly flat piece of black plastic ($n = 1.20$) as shown in Fig. 41-30*a*. They touch at A. Light of wavelength 600 nm is incident normally from above. The location of the dark fringes in the reflected light is shown on the sketch of Fig. 41-30*b*. (*a*) How thick is the space between the glass and the plastic at B? (*b*) Water ($n = 1.33$) seeps into the region between the glass and the plastic. How many dark fringes are seen when all the air has been displaced by water? (The straightness and equal spacing of the fringes is an accurate test of the flatness of the glass.)

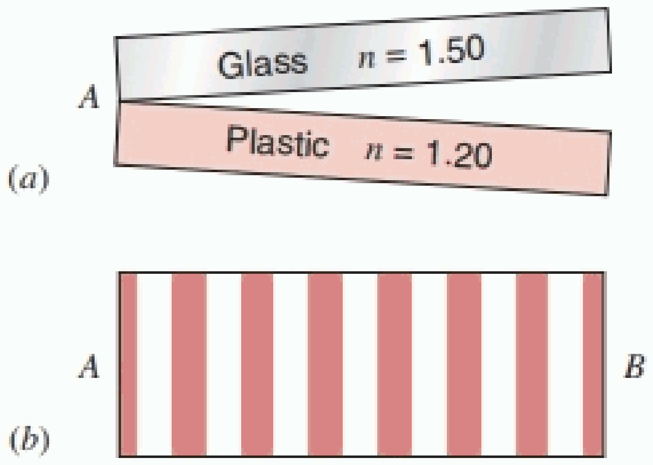


FIGURE 41-30. Problem 10.

11. In the Newton's rings experiment, show (*a*) that the difference in radius between adjacent rings (maxima) is given by

$$\Delta r = r_{m+1} - r_m \approx \tfrac{1}{2}\sqrt{\lambda R/m},$$

assuming $m \gg 1$; and (*b*) that the area between adjacent rings (maxima) is given by

$$A = \pi\lambda R,$$

assuming $m \gg 1$. Note that this area is independent of m.

12. Write an expression for the intensity observed in Michelson's interferometer (Fig. 41-20) as a function of the position of the movable mirror. Measure the position of the mirror from the point at which $d_1 = d_2$.

CHAPTER 42

DIFFRACTION

Diffraction is the bending or spreading of waves that encounter an object (a barrier or an opening) in their path. This chapter considers only diffraction of light waves, but diffraction occurs for all types of waves. Sound waves, for example, are diffracted by ordinary objects, and as a result we can hear sounds even though we may not be in a direct line to their source. For diffraction to occur, the size of the object must be of the order of the wavelength of the incident waves; when the wavelength is much smaller than the size of the object, diffraction is ordinarily not observed and the object casts a sharp shadow.

Diffraction patterns consist of light and dark bands similar to the interference patterns discussed in Chapter 41. By studying these patterns, we can learn about the diffracting object. For example, diffraction of x rays is an important method for the study of the structure of solids, and diffraction of gamma rays is used to study nuclei. Diffraction also has unwanted effects, such as the spreading of light as it enters the aperture of a telescope, which limits its ability to resolve or separate stars that appear to be close to one another. These various effects of diffraction are considered in this chapter and the following one.

42-1 DIFFRACTION AND THE WAVE THEORY OF LIGHT

When light passes through a narrow slit (of width comparable to the wavelength of the light; see Fig. 40-1), the light beams not only flare out far beyond the geometrical shadow of the slit; they also give rise to a series of alternating light and dark bands that resemble interference fringes (Fig. 42-1). In Chapter 41, we argued that the appearance of interference fringes provides strong evidence for the wave nature of light. We can also argue that the appearance of diffraction patterns similarly requires that light must travel as waves.

Although diffraction was already known at the time of Huygens and Newton, neither of them believed that it provided evidence that light must be a wave. Newton in particular believed that light traveled as a stream of particles.

A strong proponent of the wave theory of light was the French engineer Augustin Fresnel (1788–1827). Fresnel explained diffraction based on the wave theory, which was

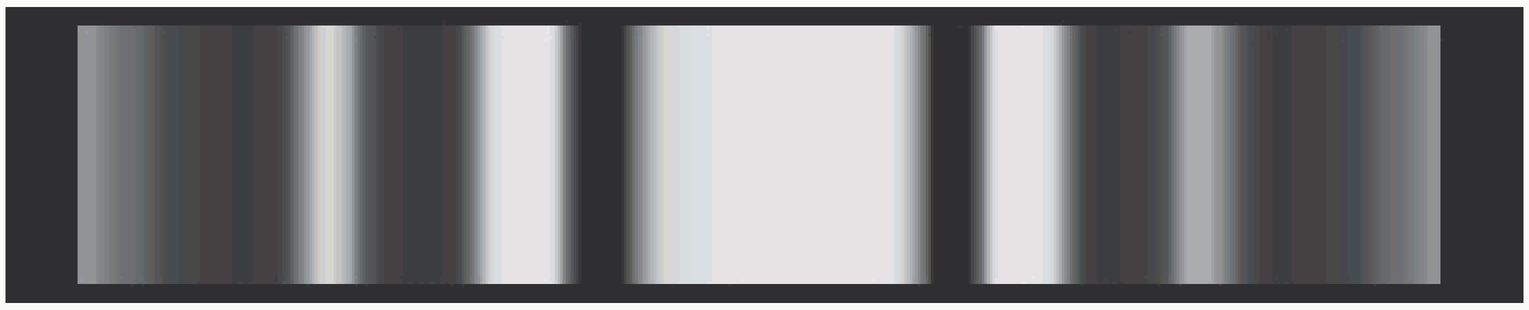


FIGURE 42-1. The diffraction pattern produced when light passes through a single narrow slit.

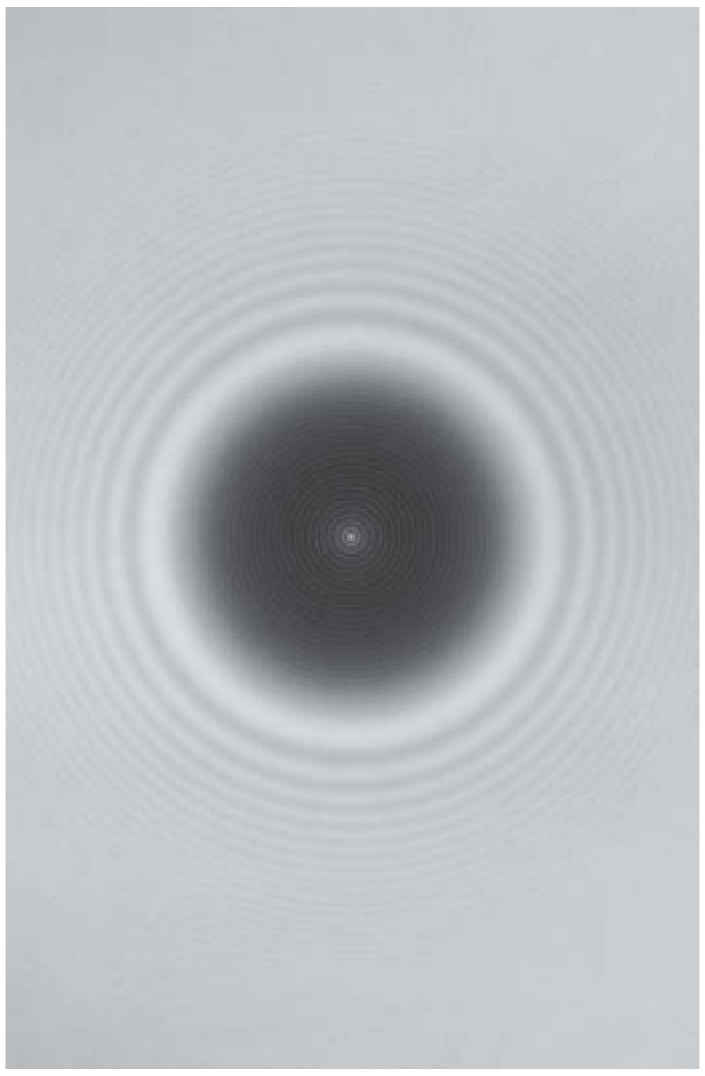


FIGURE 42-2. The diffraction pattern of a disk. Note the bright Poisson spot at the center of the pattern.

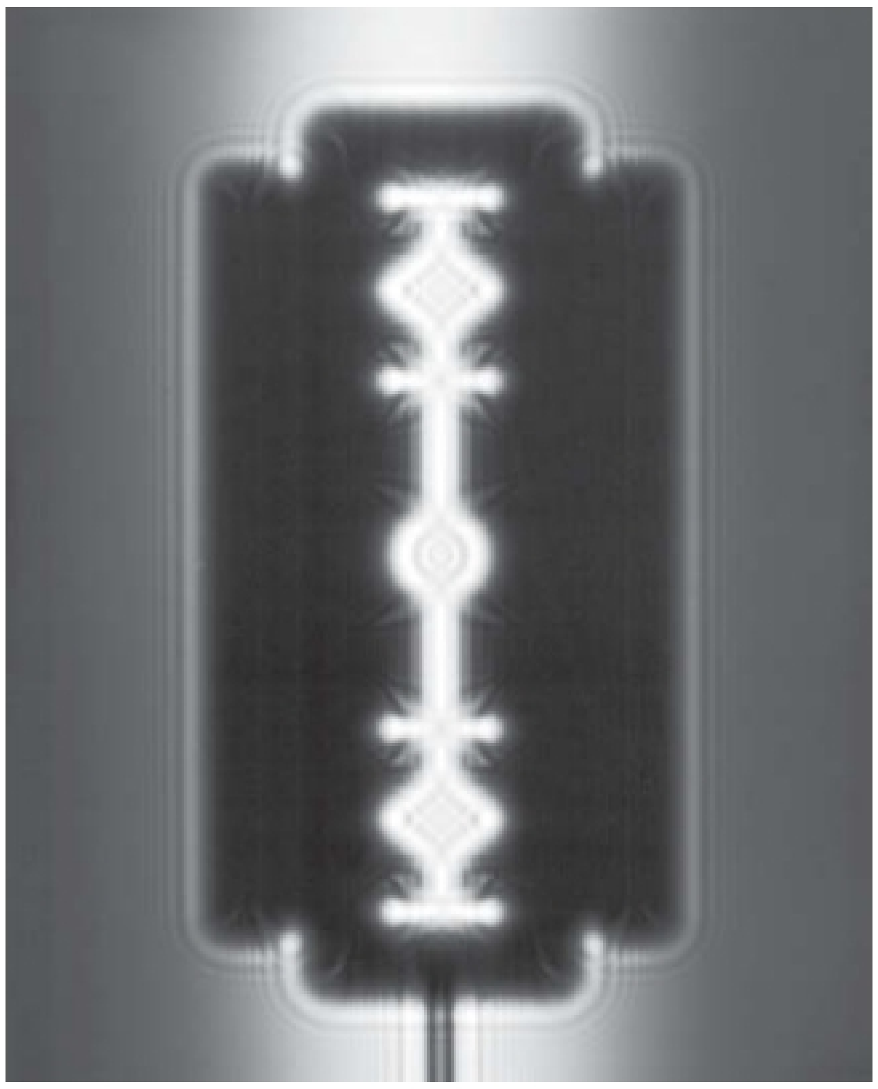


FIGURE 42-3. The diffraction pattern of a razor blade viewed in monochromatic light. Note the fringes near the edges.

not widely accepted even after Thomas Young's experiments on double-slit interference. In 1819, Fresnel submitted a paper on his theory of diffraction in a competition sponsored by the French Academy of Sciences. One of the members of the Academy, Simeon-Denis Poisson (a strong opponent of the wave theory of light), ridiculed Fresnel's theory because, as Poisson himself showed, Fresnel's diffraction theory led to the "absurd" prediction that the shadow of an opaque object should have a bright spot at its center. Figure 42-2 shows the diffraction pattern of a disk; the clearly visible bright spot at its center (known as the Poisson spot) supports Fresnel's interpretation.

Figure 42-3 shows the diffraction pattern produced when an ordinary object is illuminated by monochromatic light. Actually, you do not need special apparatus to observe diffraction. Hold two fingers so that there is a narrow slit between them, and look at a lightbulb through the slit. The dark lines you see in the slit are caused by diffraction. Another common example of diffraction is the "floaters" that many people can observe in their field of view. Floaters are translucent dots or tiny chains that appear to float and drift. They can be seen by focusing the eyes at a distance while staring at a brightly illuminated piece of white paper. Floaters are caused by blood cells and other microscopic debris in the fluid of the eyeball; what we observe is the diffraction pattern on the retina.

Figure 42-4 shows the generalized diffraction situation. The curved surfaces on the left represent wavefronts of the incident light. The light falls on the diffracting object *B*, which we show in Fig. 42-4 as an opaque barrier containing an aperture of arbitrary shape. (Later, we consider an aperture that is a single narrow slit, which produced the diffraction pattern shown in Fig. 42-1.) *C* in Fig. 42-4 is a screen or photographic film that receives the light that passes through or around the diffracting object.

We can calculate the pattern of light intensity on screen *C* by subdividing the wavefront into elementary areas $d\vec{\mathbf{A}}$, each of which becomes a source of an expanding Huygens wavelet. The light intensity at an arbitrary point *P* is found by superimposing the wave disturbances (that is, the $\vec{\mathbf{E}}$ vectors) caused by the wavelets reaching *P* from all these elementary sources.

The wave disturbances reaching *P* differ in amplitude and in phase because (1), the elementary sources are at varying distances from *P*; (2) the light leaves the elementary sources at various angles to the normal to the wave-

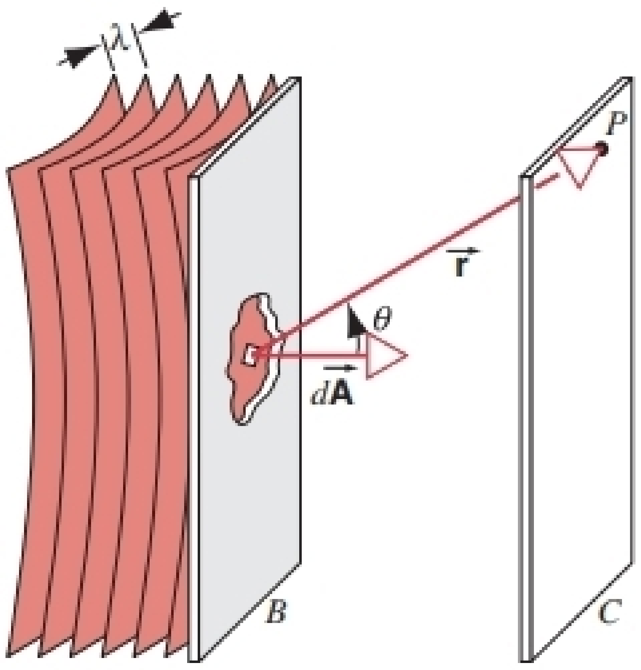


FIGURE 42-4. Diffraction occurs when coherent wavefronts of light fall on opaque barrier *B*, which contains an aperture of arbitrary shape. The diffraction pattern can be viewed on screen *C*.

front; and (3) some sources are blocked by barrier B, while others are not. Diffraction calculations, which are simple in principle, may become difficult in practice. The calculation must be repeated for every point on screen C at which we wish to know the light intensity. We followed exactly this program in calculating the double-slit intensity pattern in Section 41-4. The calculation there was simple because we assumed only two elementary sources, the two narrow slits.

Figure 42-5 shows another representation of Fig. 42-4, in the form of ray diagrams. The pattern formed on the screen depends on the separation between the screen C and the aperture B. In general, we can consider three cases:

1. *Very small separation.* When C is very close to B, the waves travel only a short distance after leaving the aperture, and the rays diverge very little. The effects of diffraction are negligible, and the pattern on the screen is the geometric shadow of the aperture.

2. *Very large separation.* Figure 42-5*a* represents the situation when the screen is so far from the aperture that we can regard the rays as parallel or, equivalently, the wavefronts as planes. (In this case, we also assume the source to be far from the aperture, so that the incident wavefronts are also planes. The same effect can be achieved by illuminating the aperture with a laser.) One way of achieving this condition, which is known as *Fraunhofer diffraction,* in the laboratory is to use two converging lenses, as in Fig. 42-5*b*. The first lens converts the diverging light from the source into a plane wave, and the second lens focuses plane waves leaving the aperture to point P. All rays that reach P leave the aperture parallel to the dashed line Px drawn from P through the center of the second lens.

3. *Intermediate separation.* In the case shown in Fig. 42-5*c*, the screen can be at any distance from the aperture, and the rays entering and leaving the aperture are not parallel. This general case is called *Fresnel diffraction.*

Although Fraunhofer diffraction is a special limiting case of the more general Fresnel diffraction, it is an important case and is easier to handle mathematically. We assumed Fraunhofer diffraction in our analysis of double-slit interference in Chapter 41. In this book we deal only with Fraunhofer diffraction.

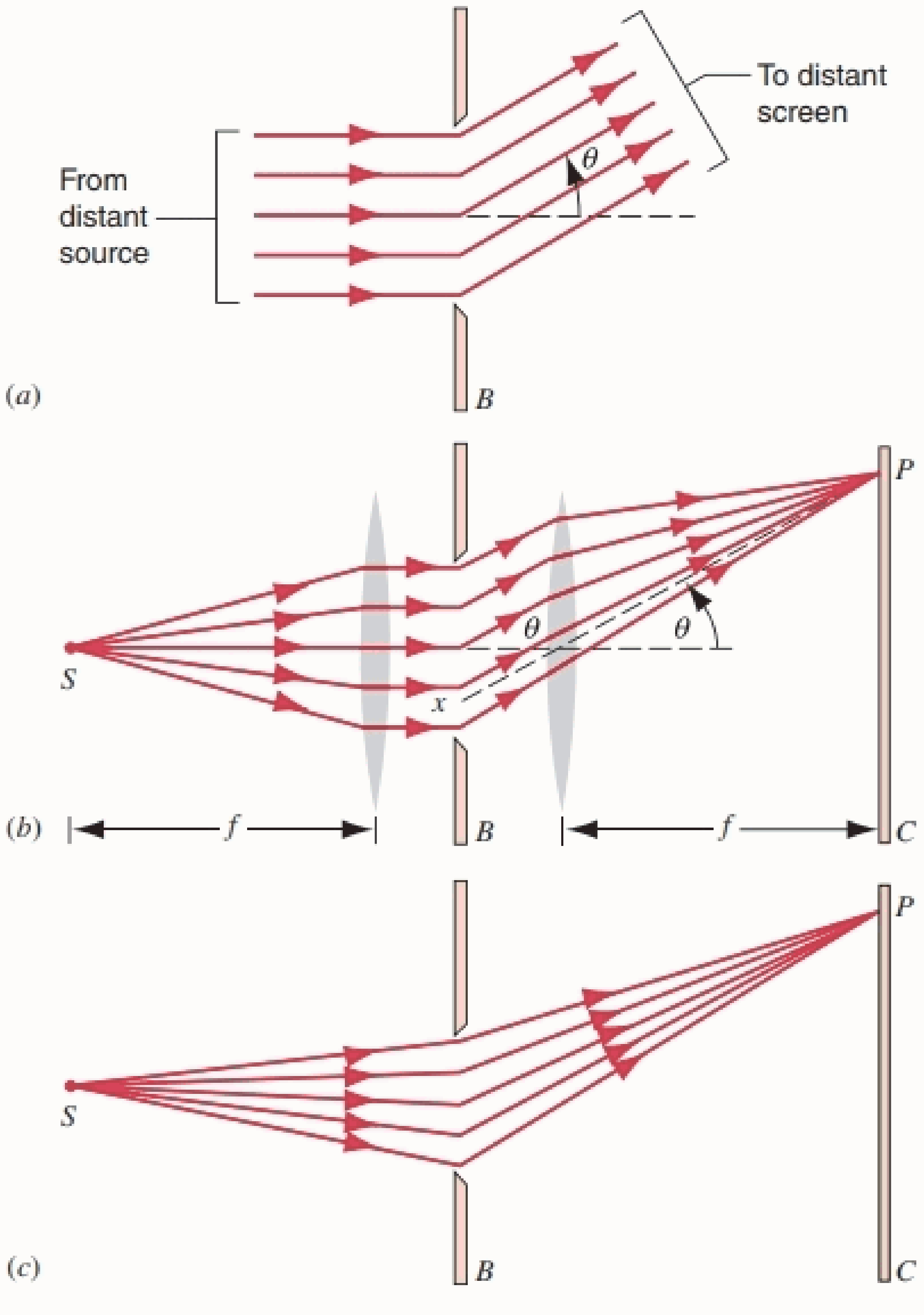


FIGURE 42-5. Light from point source S illuminates a slit in the opaque barrier B. The slit extends a long distance above and below the plane of the figure; this distance is much greater than the slit width a. The intensity at point P on screen C depends on the relative phases of the light received from various parts of the slit. (*a*) If source S and screen C are moved to large distances from the slit, both the incident and emergent light at B consist of nearly parallel rays. (*b*) Rather than using large distances, the source and the screen can each be placed in the focal plane of a lens; once again, parallel light rays enter and leave the slit. (*c*) Without the lens, the rays are not parallel.

42-2 SINGLE-SLIT DIFFRACTION

The simplest diffraction pattern to analyze is that produced by a long, narrow slit. In this section we discuss the locations of the minima and maxima in the pattern as shown in Fig. 42-1. In the next section we calculate the intensity of the pattern as a function of position on the screen.

Figure 42-6 shows a plane wave falling at normal incidence on a slit of width a. Let us first consider the central point P_0. Rays that leave the slit parallel to the central horizontal axis are brought to a focus at P_0. These rays are cer-

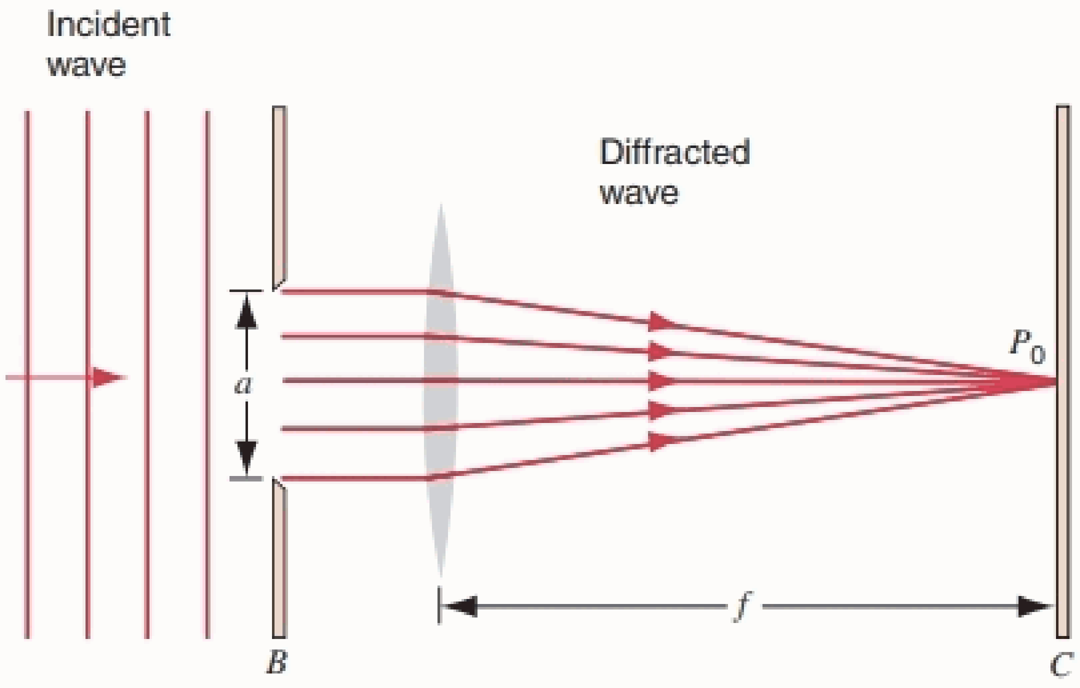


FIGURE 42-6. Conditions at the central maximum of the diffraction pattern.

tainly in phase at the plane of the slit, and they remain in phase as they are brought to a focus by the lens (see, for example, Fig. 40-21*a*). Since all rays arriving at P_0 are in phase, they interfere constructively and produce a maximum of intensity at P_0.

We now consider another point on the screen. Light rays that reach P_1 in Fig. 42-7 leave the slit at the angle θ, as shown. The ray xP_1 passes undeflected through the center of the lens and therefore determines θ. Ray r_1 originates at the top of the slit and ray r_2 at its center. If θ is chosen so that the distance bb' in the figure is one-half wavelength, r_1 and r_2 are out of phase and interfere destructively at P_1. The same is true for a ray just below r_1 and another just below r_2. In fact, for every ray passing through the upper half of the slit, there is a corresponding ray passing through the lower half, originating at a point $a/2$ below the first ray, such that the two rays are out of phase at P_1. Every ray arriving at P_1 from the upper half of the slit interferes destructively with one coming from the bottom half of the slit. The intensity at P_1 is therefore zero, and P_1 is the first minimum of the diffraction pattern.

Since the distance bb' equals $(a/2)\sin\theta$, the condition for the first minimum can be written

$$\frac{a}{2}\sin\theta = \frac{\lambda}{2},$$

or

$$a\sin\theta = \lambda. \tag{42-1}$$

Equation 42-1 shows that the central maximum becomes wider as the slit is made narrower. If the slit width is as small as one wavelength ($a = \lambda$), the first minimum occurs at $\theta = 90°$ ($\sin\theta = 1$ in Eq. 42-1), which implies that the central maximum fills the entire forward hemisphere. We assumed a condition approaching this in our discussion of double-slit interference in Section 41-2.

In Fig. 42-8, the slit is divided into four equal zones, with a ray leaving the top of each zone. Let θ be chosen so that the distance bb' is one-half wavelength. Rays r_1 and r_2 then cancel at P_2. Rays r_3 and r_4 are also one-half wavelength out of phase and also cancel. Consider four other rays, emerging from the slit a given distance below these four rays. The two rays below r_1 and r_2 cancel, as do the two rays below r_3 and r_4. We can proceed across the entire slit and conclude again that no light reaches P_2; we have located a second point of zero intensity.

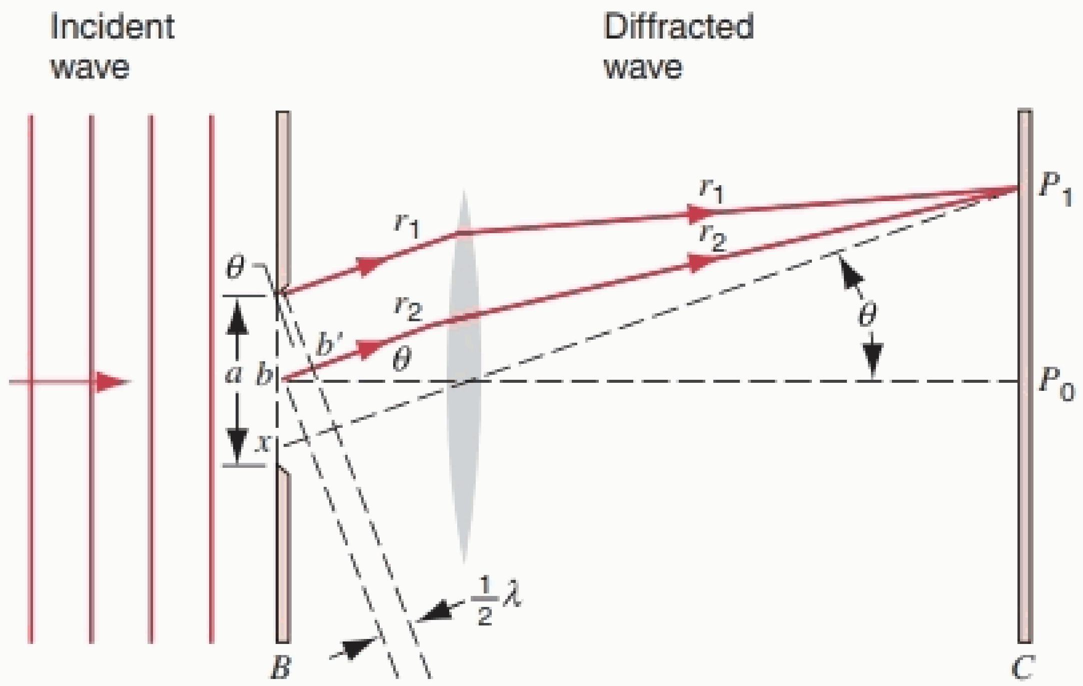


FIGURE 42-7. Conditions at the first minimum of the diffraction pattern. The angle θ is such that the distance bb' is one-half wavelength.

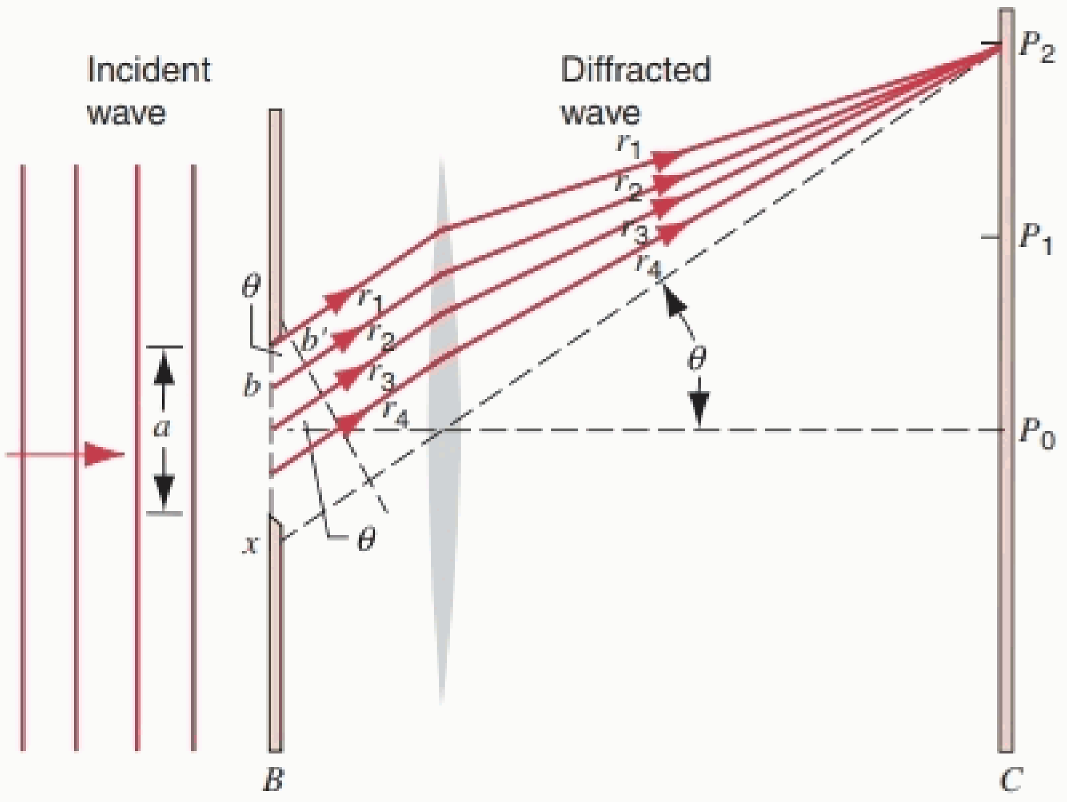


FIGURE 42-8. Conditions at the second minimum of the diffraction pattern. The angle θ is such that the distance bb' is one-half wavelength.

The condition for destructive interference requires that bb', which is $(a/4)\sin\theta$ from the geometry of Fig. 42-8, must be $\frac{1}{2}\lambda$, and so

$$\frac{a}{4}\sin\theta = \frac{\lambda}{2},$$

or

$$a\sin\theta = 2\lambda. \tag{42-2}$$

For a given slit width a and wavelength λ, Eq. 42-2 gives the position on the screen of the second minimum in terms of the angle θ. By extension of Eqs. 42-1 and 42-2, the general formula for the minima in the diffraction pattern on screen C is

$$a\sin\theta = m\lambda \qquad m = \pm1, \pm2, \pm3, \ldots \qquad \text{(minima)}. \tag{42-3}$$

There is a maximum approximately halfway between each adjacent pair of minima. Later in the chapter we derive a formula for the intensity of the diffracted light, from which the locations of the maxima can be found exactly. Note that Eq. 42-3 suggests *two* minima (and corresponding maxima) for each m, one at an angle θ above the central axis and one below (corresponding to $m < 0$). In deriving Eq. 42-3, consider how the assumption of parallel rays (Fraunhofer diffraction) has simplified the analysis.

SAMPLE PROBLEM 42-1. A slit of width a is illuminated by white light. For what value of a does the first minimum for red light ($\lambda = 650$ nm) fall at $\theta = 15°$?

Solution At the first minimum, $m = 1$ in Eq. 42-3. Solving for a, we then find

$$a = \frac{m\lambda}{\sin\theta} = \frac{(1)(650\text{ nm})}{\sin 15^\circ}$$

$$= 2510\text{ nm} = 2.51\ \mu\text{m}.$$

For the incident light to flare out that much (±15°), the slit must be very narrow indeed, amounting to about four times the wavelength (and far narrower than a fine human hair, which may only be about 100 μm in diameter).

SAMPLE PROBLEM 42-2. In Sample Problem 42-1, what is the wavelength λ' of the light whose first diffraction maximum (not counting the central maximum) falls at 15°, thus coinciding with the first minimum for red light?

Solution Maxima occur about halfway between minima, so there is a maximum at 15° when the first minimum is at 10° and the second minimum is at 20°. In this case, for the second minimum,

$$a \sin\theta = 2\lambda',$$

or

$$\lambda' = \tfrac{1}{2}(2510\text{ nm})(\sin 20^\circ) = 430\text{ nm}.$$

Light of this wavelength is violet. The second maximum for light of wavelength 430 nm always coincides with the first minimum for light of wavelength 650 nm, no matter what the slit width is. If the slit is relatively narrow, the angle θ at which this overlap occurs is relatively large, and conversely.

42-3 INTENSITY IN SINGLE-SLIT DIFFRACTION

In Section 42-2, we located the positions of the minima of the single-slit diffraction pattern. We now wish to find an expression for the intensity of the *entire* pattern as a function of the diffraction angle θ. This expression will permit us to find the location and intensity of the maxima. Before we do this calculation, let us consider some qualitative conclusions about the maxima and minima.

Figure 42-9 shows a slit of width a divided into N parallel strips, each of width δx. The strips are very narrow, so each strip can be regarded as a radiator of Huygens wavelets, and all the light from a given strip arrives at point P with the same phase. The waves arriving at P from any pair of adjacent strips have the same (constant) phase difference $\Delta\phi$, which can be found from

$$\frac{\text{phase difference}}{2\pi} = \frac{\text{path difference}}{\lambda}$$

or

$$\Delta\phi = \frac{2\pi}{\lambda}\,\delta x \sin\theta, \tag{42-4}$$

where $\delta x \sin\theta$, as the detail of Fig. 42-9 shows, is the path difference for rays originating from corresponding points of adjacent strips.

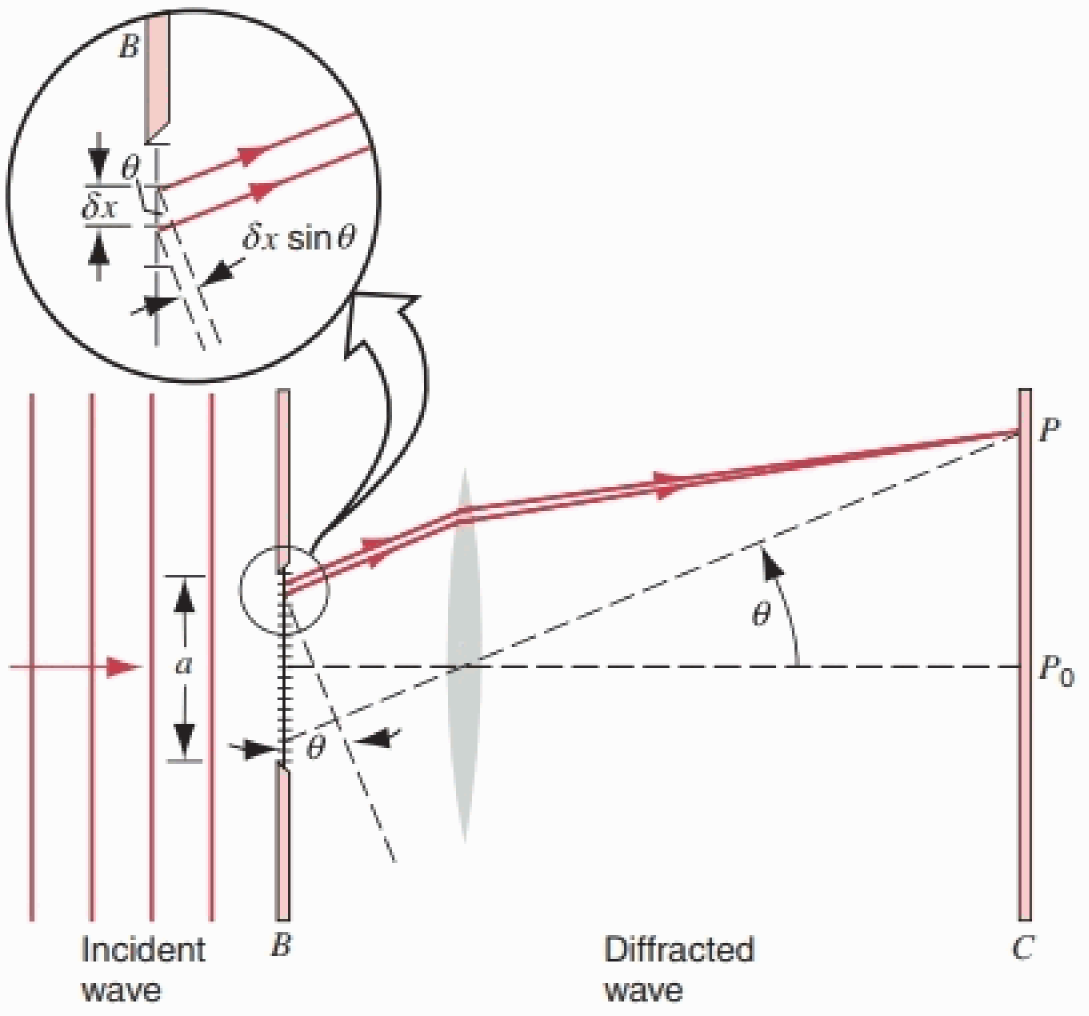


FIGURE 42-9. A slit of width a is divided into N strips of width δx. The inset shows more clearly the conditions at the second strip. In the differential limit, the width dx of each strip becomes infinitesimally small and the number of strips becomes infinitely large. Here and in the next figure we take $N = 18$ for clarity.

If the angle θ is not too large, each strip produces a wave of the same amplitude δE_0 at P. The net effect at P is due to the superposition of N vectors of the same amplitude, each differing in phase from the next by $\Delta\phi$. To find the intensity at P, we must first find the net electric field of the N vectors.

In Section 41-4, we introduced a graphical method for adding wave disturbances that enabled us to calculate the intensity in double-slit interference. That method is based on representing each wave disturbance as a phasor (a rotating vector) and finding the resultant phasor amplitude by vector addition, taking into account the relative phase given by Eq. 42-4. The resultant electric field E_θ varies with θ, because the phase difference $\Delta\phi$ varies with θ.

Let us consider some examples of the addition of phasors in single-slit diffraction. We first consider the resultant electric field at point P_0 (the center of the diffraction pattern on the screen). In this case $\theta = 0$, and Eq. 42-4 gives $\Delta\phi = 0$ as the phase difference between adjacent strips. According to the method of Section 41-4, we then lay N vectors of length δE_0 head to tail and parallel to one another ($\Delta\phi = 0$). The resultant E_θ is shown in Fig. 42-10a. This is clearly the maximum value that the resultant of these N vectors can take, so we label it E_m.

As we move away from $\theta = 0$, the phase difference $\Delta\phi$ assumes a definite nonzero value. Again laying the N vectors head to tail, each differing in direction from the previous one by $\Delta\phi$, we obtain the resultant shown in Fig. 42-10b. Note that E_θ is smaller than it was in Fig. 42-10a.

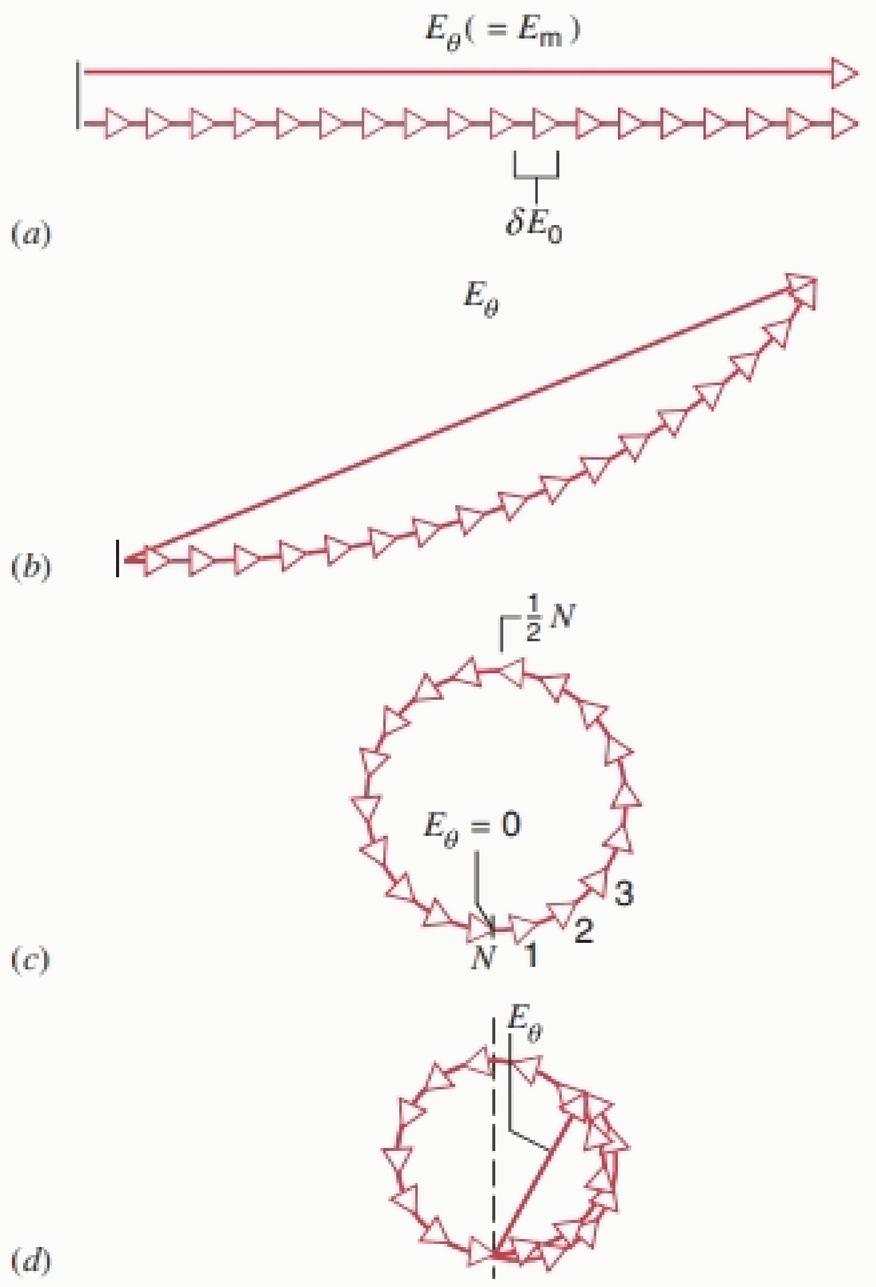


FIGURE 42-10. Phasors in single-slit diffraction, showing conditions at (*a*) the central maximum, (*b*) a direction slightly removed from the central maximum, (*c*) the first minimum, and (*d*) the first maximum beyond the central maximum. This figure corresponds to $N = 18$ in Fig. 42-9.

Now consider the first minimum of the diffraction pattern (point P_1 in Fig. 42-7). At this point the intensity is zero, so the resultant E_θ must be zero. This means that the N phasors, laid head to tail, must form a closed loop, as in Fig. 42-10*c*.

Beyond the first minimum, the phase shift $\Delta\phi$ is still larger, and the chain of vectors coils around through an angle greater than 360°. At a certain angle (corresponding to a certain phase shift, as in Fig. 42-10*d*), the resultant E_θ has its greatest length within this loop, corresponding to the first maximum beyond the central one. Note that the intensity of this maximum is much smaller than the intensity of the central maximum, represented in Fig. 42-10*a*. Eventually, this loop closes on itself, giving a resultant of zero and corresponding to the second minimum.

Calculating the Intensity

Now we turn to a full calculation of the intensity at any location θ on the screen. Our method in finding the intensity of the single-slit diffraction pattern for any θ is to evaluate the phase shift according to Eq. 42-4 and find the resultant E_θ, as in Fig. 42-10*b*. The square of this resultant then gives the relative intensity, as in Section 41-4.

The light arriving at P from a given strip is in phase only if the strip is infinitesimally small and the number of strips is correspondingly large. The chain of phasors of Fig. 42-10*b* then approaches the arc of a circle, as drawn in Fig. 42-11. The length of the arc is E_m, and the amplitude we seek for the resultant field is indicated by the chord E_θ. The angle ϕ is the total phase difference between the rays from the top and bottom of the strip; as Fig. 42-11 shows, ϕ is also the angle between the two radii R.

From this figure we can write

$$E_\theta = 2R \sin\frac{\phi}{2}.$$

From Fig. 42-11, ϕ in radian measure is

$$\phi = \frac{E_m}{R}.$$

Combining to eliminate R yields

$$E_\theta = \frac{E_m}{\phi/2} \sin\frac{\phi}{2}$$

or

$$E_\theta = E_m \frac{\sin\alpha}{\alpha}, \tag{42-5}$$

in which

$$\alpha = \frac{\phi}{2}. \tag{42-6}$$

From Fig. 42-9, recalling that ϕ is the phase difference between rays from the top and the bottom of the slit and that the path difference for these rays is $a \sin\theta$, we have

$$\frac{\text{phase difference}}{2\pi} = \frac{\text{path difference}}{\lambda},$$

or

$$\phi = \frac{2\pi}{\lambda}(a \sin\theta).$$

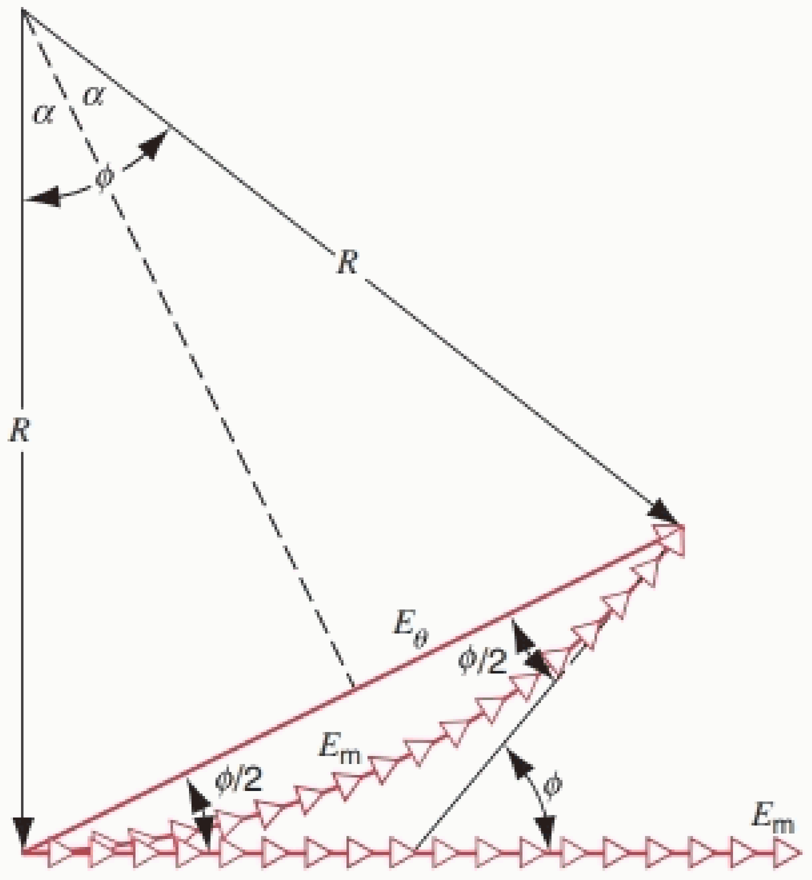


FIGURE 42-11. A construction used to calculate the intensity in single-slit diffraction. The situation corresponds to that of Fig. 42-10*b*.

Combining with Eq. 42-6 yields

$$\alpha = \frac{\phi}{2} = \frac{\pi a}{\lambda} \sin \theta. \tag{42-7}$$

Equation 42-5, with α evaluated according to Eq. 42-7, gives the amplitude of the wave disturbance for a single-slit diffraction pattern at any angle θ. The intensity I_θ for the pattern is proportional to the square of the amplitude, so

$$I_\theta = I_{\mathrm{m}} \left(\frac{\sin \alpha}{\alpha} \right)^2. \tag{42-8}$$

Equation 42-8, combined with Eq. 42-7, gives the result we seek for the intensity of the single-slit diffraction pattern at any θ. Figure 42-12 shows plots of the relative intensity I_θ/I_{m} for several values of the ratio a/λ. Note that the pattern becomes narrower as we increase a/λ.

Minima occur in Eq. 42-8 when $\sin \alpha = 0$, or

$$\alpha = m\pi \qquad m = \pm 1, \pm 2, \pm 3, \ldots . \tag{42-9}$$

Combining with Eq. 42-7 leads to

$$a \sin \theta = m\lambda \qquad m = \pm 1, \pm 2, \pm 3, \ldots \qquad \text{(minima)},$$

which is the result derived in the preceding section (Eq. 42-3). In that section, however, we derived *only* this result, obtaining no quantitative information about the intensity of the diffraction pattern at places in which it was not zero. Here (Eq. 42-8) we have complete intensity information.

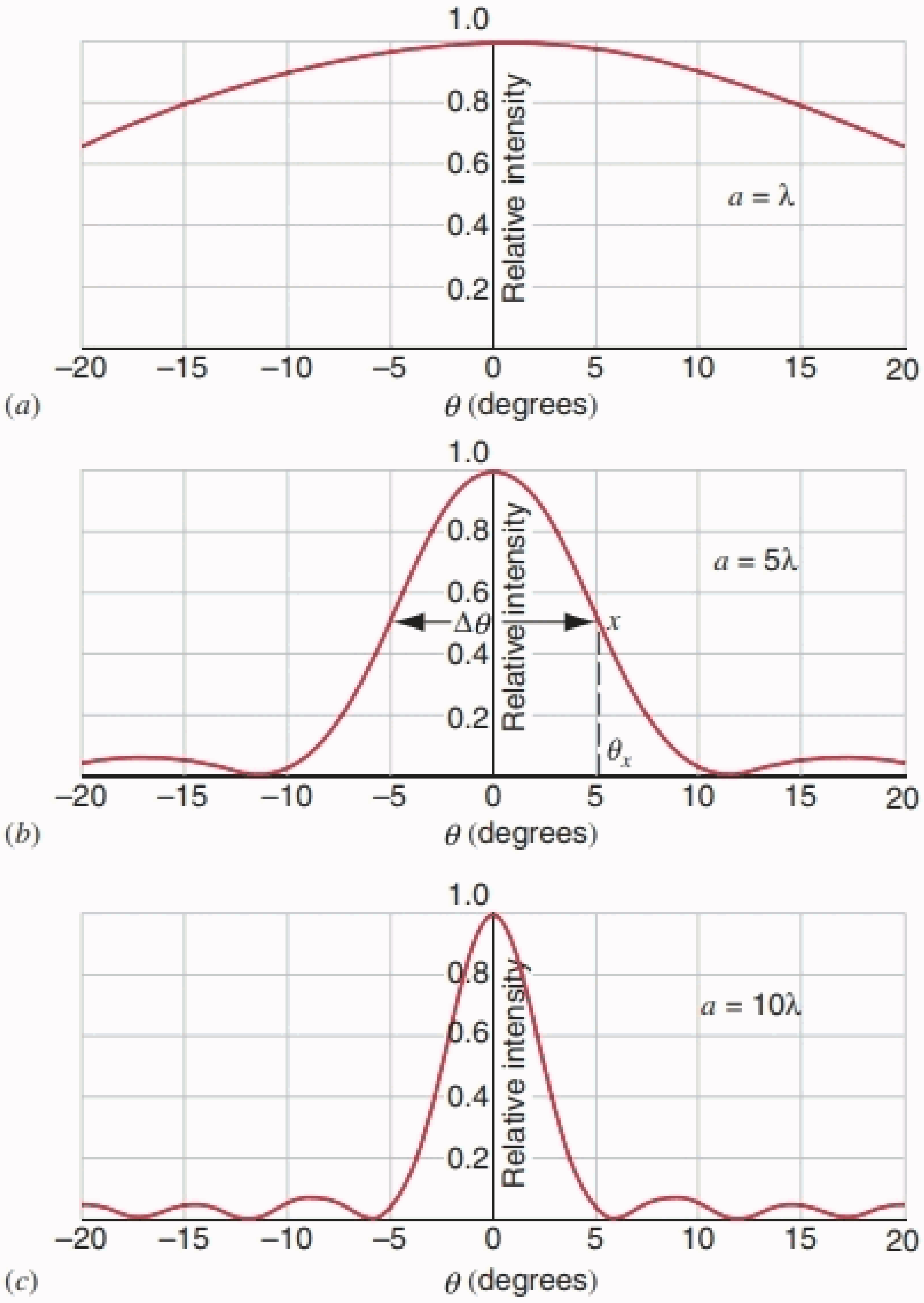


FIGURE 42-12. The intensity distribution in single-slit diffraction for three different values of the ratio a/λ. The wider the slit, the narrower is the central diffraction peak. As indicated in (*b*), $\Delta\theta$ gives a measure of the width of the central peak.

SAMPLE PROBLEM 42-3. Calculate, approximately, the relative intensities of the maxima in the single-slit Fraunhofer diffraction pattern.

Solution The maxima lie approximately halfway between the minima and are roughly given by (see Problem 3)

$$\alpha \approx (m + \tfrac{1}{2})\pi \qquad m = 1, 2, 3, \ldots ,$$

with a similar result for $m < 0$. Substituting into Eq. 42-8 yields

$$I_\theta = I_{\mathrm{m}} \left[\frac{\sin (m + \frac{1}{2})\pi}{(m + \frac{1}{2})\pi} \right]^2,$$

which reduces to

$$\frac{I_\theta}{I_{\mathrm{m}}} = \frac{1}{(m + \frac{1}{2})^2 \pi^2} .$$

This yields $I_\theta/I_{\mathrm{m}} = 0.045$ ($m = 1$), 0.016 ($m = 2$), 0.0083 ($m = 3$), and so forth. The successive maxima decrease rapidly in intensity.

SAMPLE PROBLEM 42-4. Find the width $\Delta\theta$ of the central maximum in a single-slit Fraunhofer diffraction (see Fig. 42-12*b*). The width can be represented as the angle between the two points in the pattern where the intensity is one-half that at the center of the pattern.

Solution Point x in Fig. 42-12*b* is chosen so that $I_\theta = \frac{1}{2} I_{\mathrm{m}}$, or, from Eq. 42-8,

$$\frac{1}{2} = \left(\frac{\sin \alpha_x}{\alpha_x} \right)^2 .$$

This equation cannot be solved analytically for α_x. Using a pocket calculator or a computer, we can find an approximate solution to any desired accuracy. Let us rewrite the equation as

$$\alpha_x = \sqrt{2} \sin \alpha_x . \tag{42-10}$$

To solve this on your calculator, enter the "radian" mode. Pick any starting value for α_x—say, $\alpha_x = 1$. Plug this value into the right side of Eq. 42-10 and solve, obtaining 1.19. Equation 42-10 requires that this value must then be equal to α_x, which it is clearly not ($1 \neq 1.19$). Take 1.19 as the new trial value, and again evaluate the right-hand side, obtaining 1.31. We still do not have a solution that satisfies Eq. 42-10 ($1.19 \neq 1.31$), but we are closer than we were on our first try. Continue in this way, using the result of one calculation as the starting point of the next, until the difference between the calculated value of the right-hand side of Eq. 42-10 and the starting value becomes as small as you like. (You can set this up as a program for a calculator or a computer and have it do the repetitions automatically. You can also have it stop when the difference between successive values becomes smaller than a limit you can set.) This method is called the *iterative* technique for solving equations. After 10 iterations, the result is

$$\alpha_x = 1.39156,$$

and additional iterations change only the fifth decimal place. Inserting this value into Eq. 42-7 (and putting $a = 5\lambda$, as in Fig.

42-12*b*), we obtain

$$\theta_x = \sin^{-1}\left(\frac{\alpha_x \lambda}{\pi a}\right) = \sin^{-1}\left(\frac{1.39}{5\pi}\right) = 5.1°.$$

The width of the curve is then found from

$$\Delta\theta = 2\theta_x = 10.2°.$$

42-4 DIFFRACTION AT A CIRCULAR APERTURE

In focusing an image, a lens passes only the light that falls within its circular perimeter. From this point of view, a lens behaves like a circular aperture in an opaque screen. Such an aperture forms a diffraction pattern analogous to that of a single slit. Diffraction effects often limit the ability of telescopes and other optical instruments to form precise images.

The image formed by a lens can be distorted by other effects, including chromatic and spherical aberrations. These effects can be substantially reduced or eliminated by suitable shaping of the lens surfaces or by introducing correcting elements into the optical system. However, no amount of clever design can eliminate the effects of diffraction, which are determined only by the size of the aperture (the diameter of the lens) and the wavelength of the light. In diffraction, nature imposes a fundamental limitation on the precision of our instruments.

When we used geometrical optics to analyze lenses, we assumed diffraction not to occur. However, geometrical optics is itself an approximation, being the limit of wave optics. If we were to make a rigorous wave-optical analysis of the formation of an image by a lens, we would find that diffraction effects arise in a natural way.

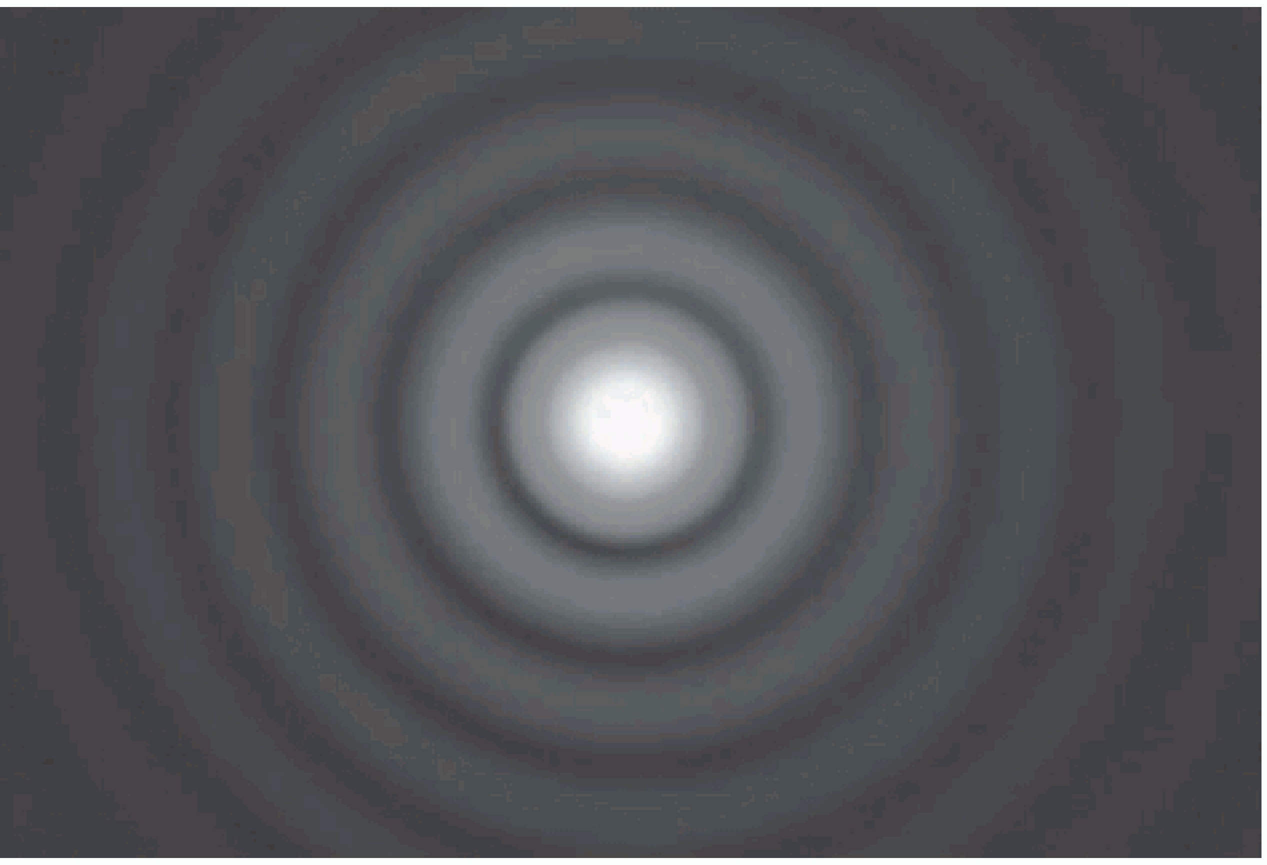


FIGURE 42-13. The diffraction pattern of a circular aperture. The central maximum is sometimes called the Airy disk (after Sir George Airy, who first solved the problem of diffraction by a circular aperture in 1835). Note the circular secondary maxima.

Figure 42-13 shows the image of a distant point source of light (a star) formed on a photographic film placed in the focal plane of the converging lens of a telescope. It is not a point, as the (approximate) geometrical optics treatment suggests, but a circular disk surrounded by several progressively fainter secondary rings. Comparison with Fig. 42-1 leaves little doubt that we are dealing with a diffraction phenomenon.

The mathematical analysis of diffraction by a circular aperture, which is beyond the level of this text, shows that (under Fraunhofer conditions) the first minimum occurs at an angle from the central axis given by

$$\sin\theta = 1.22\frac{\lambda}{d}, \qquad (42\text{-}11)$$

where d is the diameter of the aperture. This is to be compared with Eq. 42-1,

$$\sin\theta = \frac{\lambda}{a},$$

which locates the first minimum of a slit of width a. These expressions differ by the factor 1.22, which arises when we divide the circular aperture into elementary Huygens sources and integrate over the aperture.

The fact that lens images are diffraction patterns is important when we wish to distinguish two distant point objects whose angular separation is small. Figure 42-14 shows the visual appearances and the corresponding intensity patterns for two distant point objects (stars, say) with small angular separations and approximately equal central intensities. In Fig. 42-14*a* the objects are not resolved; that is, they cannot be distinguished from a single point object. In Fig. 42-14*b* they are barely resolved, and in Fig. 42-14*c* they are fully resolved.

In Fig. 42-14*b* the angular separation of the two point sources is such that the central maximum of the diffraction pattern of one source falls on the first minimum of the diffraction pattern of the other. This is called *Rayleigh's criterion* for resolving images. From Eq. 42-11, two objects that are barely resolvable by Rayleigh's criterion must have an angular separation θ_R of

$$\theta_R = \sin^{-1}\left(\frac{1.22\lambda}{d}\right).$$

Since the angles involved are rather small, we can replace $\sin\theta_R$ by θ_R, so

$$\theta_R = 1.22\frac{\lambda}{d}, \qquad (42\text{-}12)$$

in which θ_R is expressed in radians. If the angular separation θ between the objects is greater than θ_R, we can resolve the two objects; if it is less, we cannot. The angle θ_R is the smallest angular separation for which resolution is possible, using Rayleigh's criterion.

When we wish to use a lens to resolve objects of small angular separation, it is desirable to make the central disk of the diffraction pattern as small as possible. This can be done (see Eq. 42-12) by increasing the lens diameter or by

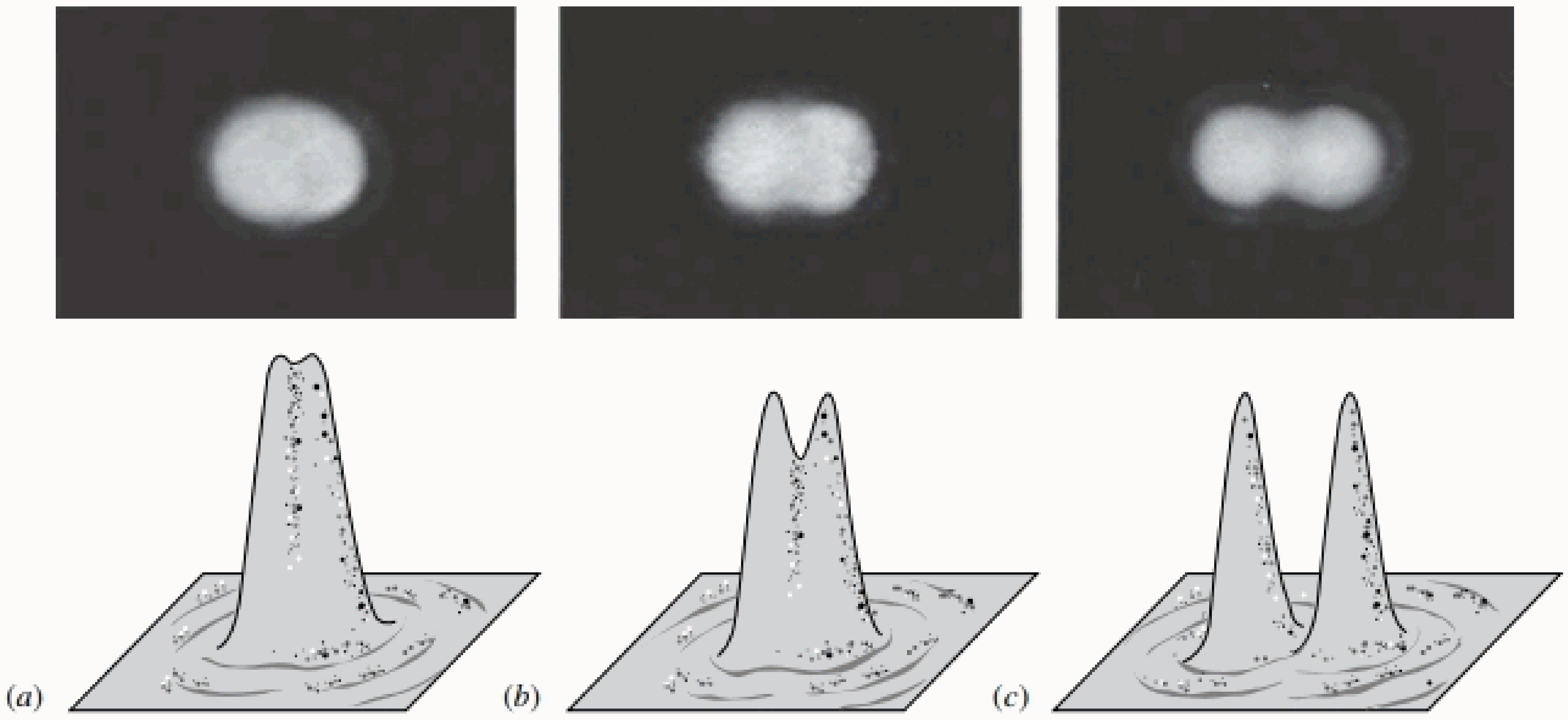


FIGURE 42-14. The images of two distant point sources (stars) formed by a converging lens. The diameter of the lens (which is the diffracting aperture) is 10 cm, so that $a/\lambda = 200{,}000$ if the effective wavelength is about 500 nm. In (*a*) the stars are so close together that their images can scarcely be distinguished, due to the overlap of their diffraction patterns. In (*b*) the stars are farther apart and their separation meets Rayleigh's criterion for resolution of their images. In (*c*) the stars are still farther apart and their images are well resolved. Plots of the intensities are shown below the images.

using a shorter wavelength. One reason for constructing large telescopes is to produce *sharper* images so that we can examine astronomical objects in finer detail. The images are also *brighter,* not only because the energy is concentrated into a smaller diffraction disk but because the larger lens collects more light. Thus fainter objects—for example, distant galaxies—can be seen.

To reduce diffraction effects in *microscopes* we often use ultraviolet light, which, because of its shorter wavelength, permits finer detail to be examined than would be possible if the same microscope used visible light. We shall see in Chapter 46 that beams of electrons behave like waves under some circumstances. In the *electron microscope* such beams may have an effective wavelength of 4×10^{-3} nm, of the order of 10^5 times shorter than that of visible light. This permits the detailed examination of tiny objects such as bacteria or viruses (Fig. 42-15). If such a small object were examined with an optical microscope, its structure would be hopelessly concealed by diffraction.

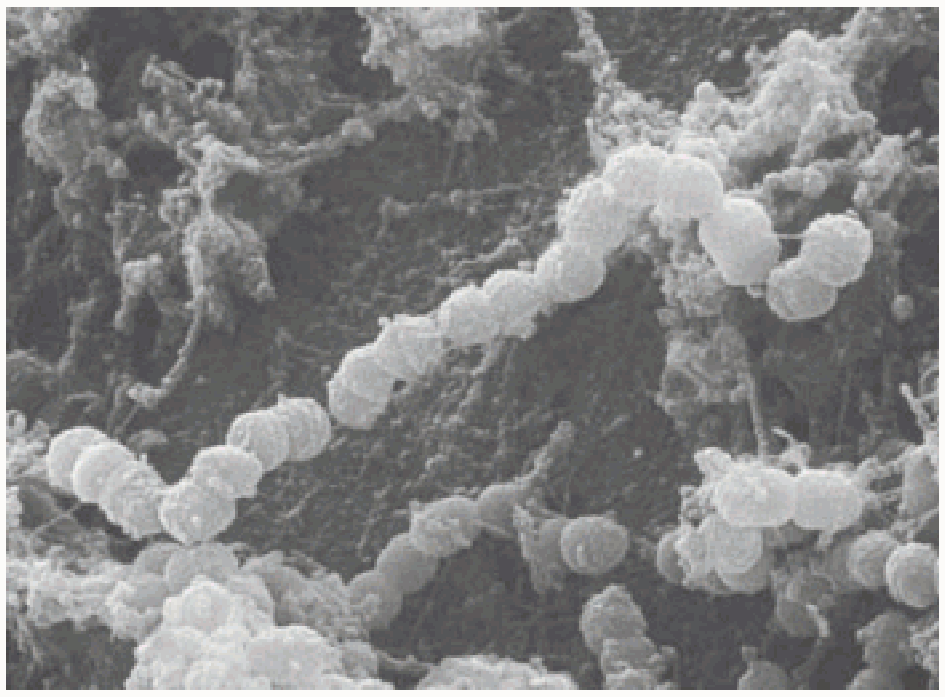


FIGURE 42-15. An image of a chain of streptococcus bacteria (diameter 10^{-6} m) obtained with an electron microscope. Note the sharpness of the image, which would not be possible using visible light.

SAMPLE PROBLEM 42-5. A converging lens 32 mm in diameter has a focal length f of 24 cm. (*a*) What angular separation must two distant point objects have to satisfy Rayleigh's criterion? Assume that $\lambda = 550$ nm. (*b*) How far apart are the centers of the diffraction patterns in the focal plane of the lens?

Solution (*a*) From Eq. 42-12,

$$\theta_R = 1.22\,\frac{\lambda}{d} = \frac{(1.22)(550 \times 10^{-9}\ \text{m})}{32 \times 10^{-3}\ \text{m}}$$
$$= 2.10 \times 10^{-5}\ \text{rad} = 4.3\ \text{arc seconds}.$$

(*b*) The linear separation is

$$\Delta x = f\theta_R = (0.24\ \text{m})(2.10 \times 10^{-5}\ \text{rad})$$
$$= 5.0\ \mu\text{m},$$

or about 9 wavelengths of the light.

42-5 DOUBLE-SLIT INTERFERENCE AND DIFFRACTION COMBINED

In our analysis of double-slit interference (Section 41-2) we assumed that the slits were arbitrarily narrow—that is, that $a \ll \lambda$. For such narrow slits, the central part of the screen on which the light falls is uniformly illuminated by the diffracted waves from each slit. When such waves interfere, they produce interference fringes of uniform intensity.

In practice, for visible light, the condition $a \ll \lambda$ is usually not met. For such relatively wide slits, the intensity of the interference fringes formed on the screen is *not* uniform. Instead, the intensity of the fringes varies within an envelope due to the diffraction pattern of a single slit.

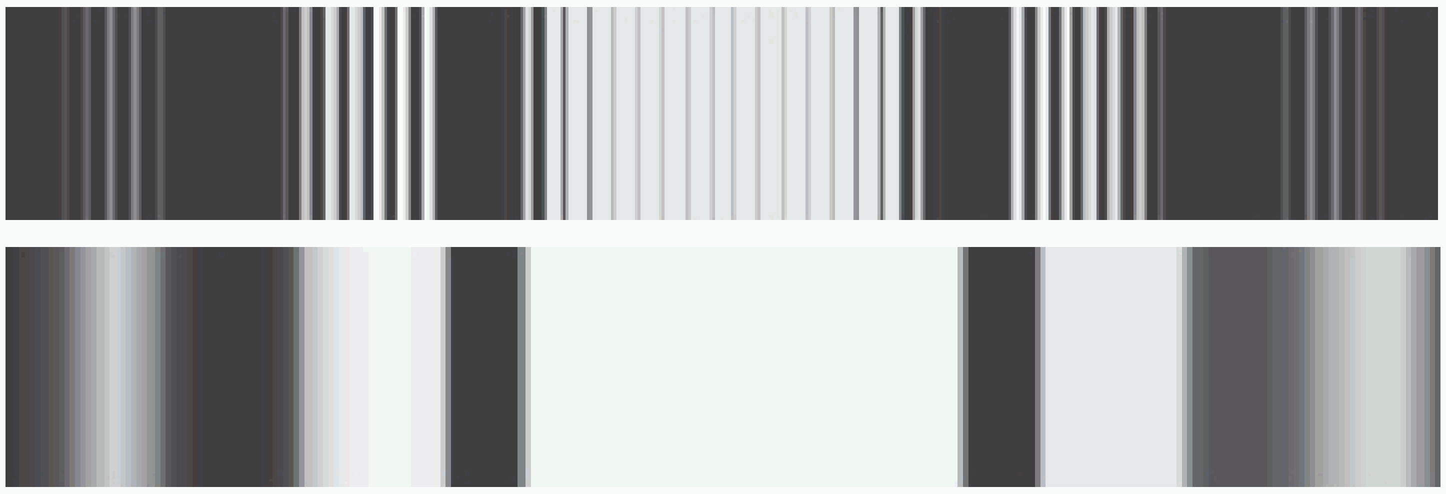


FIGURE 42-16. (*a*) Interference fringes for a double-slit system in which the slit width is not negligible in comparison with the wavelength. (*b*) The diffraction pattern of a single slit of the same width. Note that the diffraction pattern modulates the intensity of the interference fringes, as shown in part (*a*).

The effect of diffraction on a double-slit interference pattern is illustrated in Fig. 42-16, which compares the double-slit pattern with the diffraction pattern produced by a single slit of the same width as each of the double slits. You can see from Fig. 42-16*a* that the diffraction does indeed provide an intensity envelope for the more closely spaced double-slit interference fringes.

Let us now analyze the combined interference and diffraction pattern of Fig. 42-16*a*. The interference pattern for two infinitesimally narrow slits is given by Eq. 41-13 or, with a small change in notation,

$$I_{\theta,\text{int}} = I_{\text{m,int}} \cos^2 \beta, \tag{42-13}$$

where

$$\beta = \frac{\pi d}{\lambda} \sin \theta, \tag{42-14}$$

in which d is the distance between the centerlines of the slits.

The intensity for the diffracted wave from either slit is given by Eq. 42-8 or, again with a small change in notation,

$$I_{\theta,\text{dif}} = I_{\text{m,dif}} \left(\frac{\sin \alpha}{\alpha} \right)^2, \tag{42-15}$$

where

$$\alpha = \frac{\pi a}{\lambda} \sin \theta. \tag{42-16}$$

We find the combined effect by regarding $I_{\text{m,int}}$ in Eq. 42-13 as a variable amplitude, given in fact by $I_{\theta,\text{dif}}$ of Eq. 42-15. This assumption, for the combined pattern, leads to

$$I_\theta = I_{\text{m}} (\cos \beta)^2 \left(\frac{\sin \alpha}{\alpha} \right)^2, \tag{42-17}$$

in which we have dropped all subscripts referring separately to interference and diffraction. Later in this section we derive this result using phasors.

Let us express this result in words. At any point on the screen the available light intensity from each slit, considered separately, is given by the diffraction pattern of that slit (Eq. 42-15). The diffraction patterns for the two slits, again considered separately, coincide because parallel rays in Fraunhofer diffraction are focused at the same spot. Because the two diffracted waves are coherent, they interfere.

The effect of interference is to redistribute the available energy over the screen, producing a set of fringes. In Section 41-2, where we assumed $a \ll \lambda$, the available energy was virtually the same at all points on the screen so that the interference fringes had virtually the same intensities (see Fig. 41-9). If we relax the assumption that $a \ll \lambda$, the available energy is *not* uniform over the screen but is given by the diffraction pattern of a slit of width a. In this case the interference fringes have intensities that are determined by the intensity of the diffraction pattern at the location of a particular fringe. Equation 42-17 is the mathematical expression of this argument. This is especially clear in Fig. 42-17, which shows (*a*) the "interference factor" in Eq. 42-

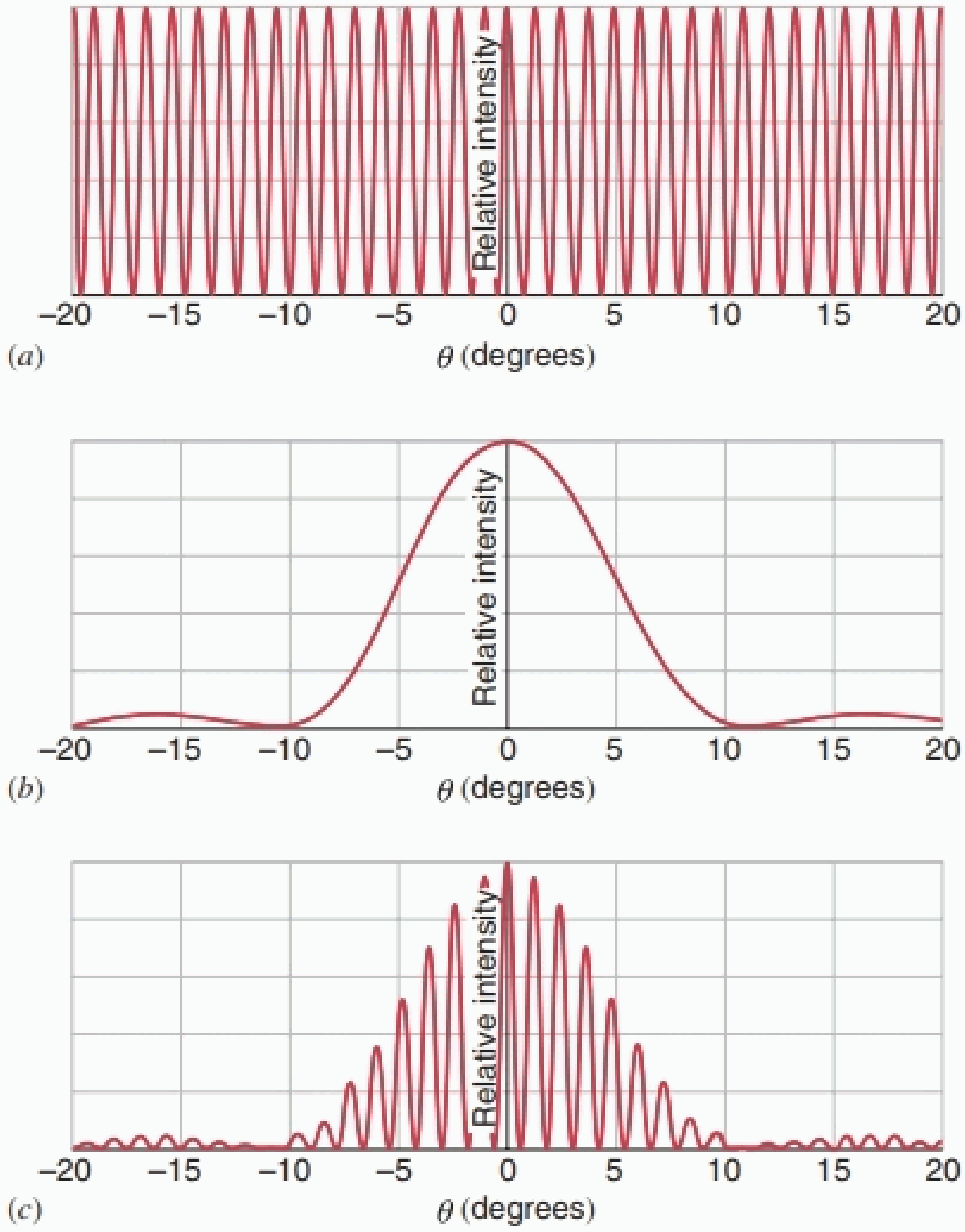


FIGURE 42-17. (*a*) Interference fringes that would be produced by a double slit of vanishingly narrow widths. (*b*) The diffraction pattern for a slit of finite width. (*c*) The pattern of interference fringes formed by two slits of the same width as that of (*b*). This pattern is equivalent to the product of the curves shown in (*a*) and (*b*). Compare Fig. 42-16*a*.

17 (that is, the factor $\cos^2 \beta$), (*b*) the "diffraction factor" $(\sin \alpha/\alpha)^2$, and (*c*) their product.

Figure 42-18 is a plot of the relative intensity I_θ/I_m given by Eq. 42-17 for $d = 50\lambda$ and for three values of a/λ. It shows clearly that for narrow slits ($a = \lambda$) the fringes are nearly uniform in intensity. As the slits are widened, the intensities of the fringes are markedly modulated by the "diffraction factor" in Eq. 42-17—that is, by the factor $(\sin \alpha/\alpha)^2$; compare with Fig. 42-12.

If we decrease the slit width a, the envelope of the fringe pattern becomes broader, and the central peak spreads out (compare Figs. 42-18*a* and 42-18*b*). As the slit width a approaches zero, $\alpha \rightarrow 0$ and $\sin \alpha/\alpha \rightarrow 1$. Thus Eq. 42-17 reduces to Eq. 42-13, which describes interference from a pair of vanishingly narrow slits. If we let the slit separation d approach zero, the two slits coalesce into a single slit of width a. From Eq. 42-14, $\beta \rightarrow 0$ as $d \rightarrow 0$, and Eq. 42-17 reduces to Eq. 42-15, the diffraction equation for a single slit of width a.

If we increase the slit width a, the envelope of the fringe pattern becomes narrower, and the central peak becomes sharper (compare Figs. 42-18*b* and 42-18*c*). The separation between the fringes, which depends on d/λ, does not change. If we increase the slit separation d, the fringes are closer together, but the envelope of the fringe pattern, which depends on a/λ, does not change.

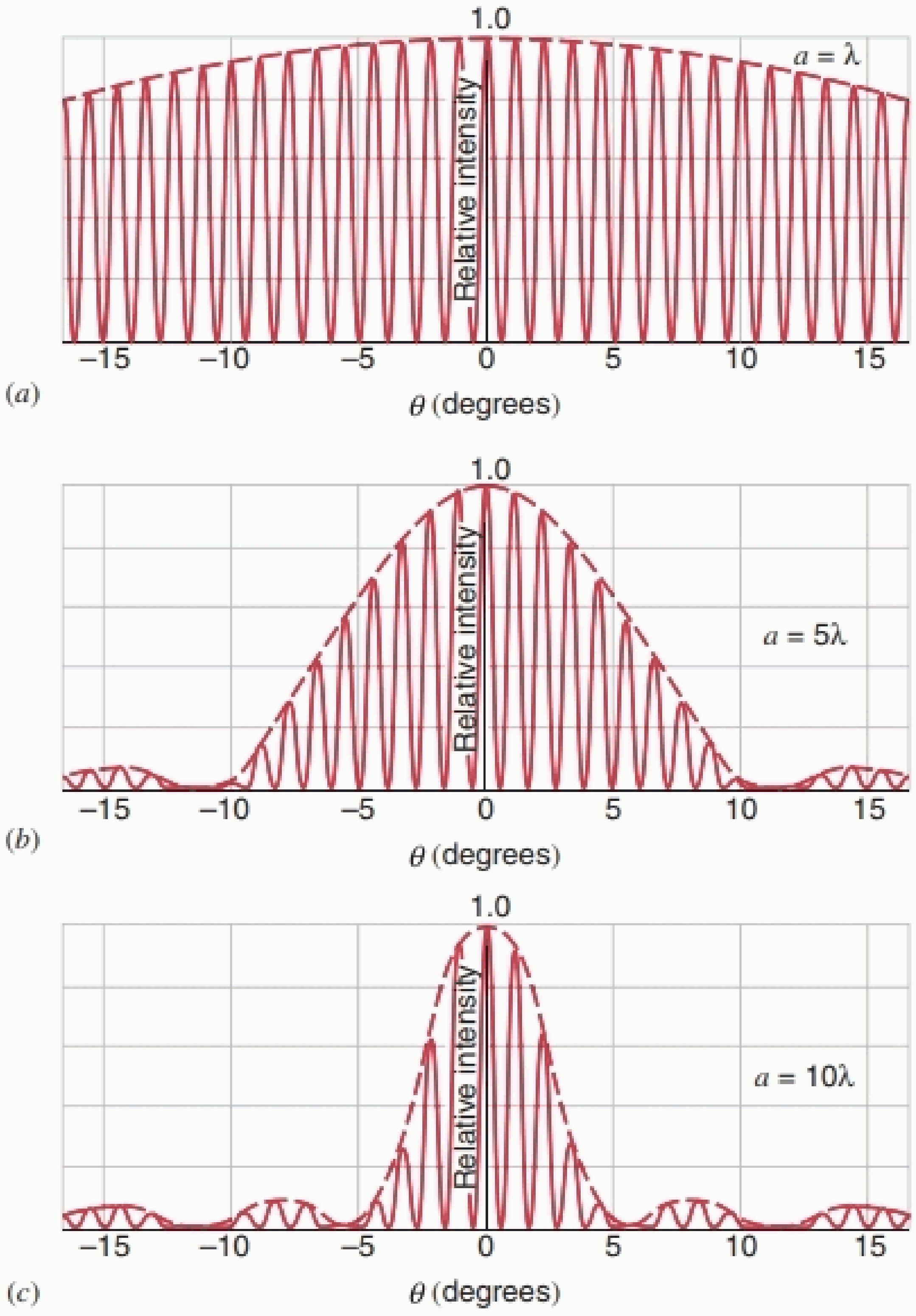


FIGURE 42-18. Interference fringes for a double slit with slit separation $d = 50\lambda$. Three different slit widths are shown.

If we increase the wavelength of the incident light, both the diffraction and interference patterns broaden: the diffraction envelope becomes wider and the fringe separation increases. The reverse effect occurs as we decrease the wavelength. Put another way, the relationship between the diffraction envelope and the interference fringes (for example, the number of fringes in the central peak) depends on the ratio d/a and is independent of λ.

The double-slit pattern illustrated in Fig. 42-17 combines interference and diffraction in an intimate way. Both are superposition effects that depend on adding wave disturbances at a given point, taking phase differences properly into account. If the waves to be combined originate from a *finite* (and usually small) number of elementary coherent radiators, as in the double slit, we call the effect *interference*. If the waves to be combined originate by subdividing a wave into *infinitesimal* coherent radiators, as in our treatment of a single slit, we call the effect *diffraction*. This distinction between interference and diffraction is convenient and useful. However, it should not cause us to lose sight of the fact that both are superposition effects and that often both are present simultaneously, as in the double-slit experiment.

SAMPLE PROBLEM 42-6. In a double-slit experiment, the distance D of the screen from the slits is 52 cm, the wavelength λ is 480 nm, the slit separation d is 0.12 mm, and the slit width a is 0.025 mm. (*a*) What is the spacing between adjacent fringes? (*b*) What is the distance from the central maximum to the first minimum of the fringe envelope?

Solution (*a*) The intensity pattern is given by Eq. 42-17, the fringe spacing being determined by the interference factor $\cos^2 \beta$. From Eq. 41-4, we have

$$\Delta y = \frac{\lambda D}{d} = \frac{(480 \times 10^{-9}\,\mathrm{m})(52 \times 10^{-2}\,\mathrm{m})}{0.12 \times 10^{-3}\,\mathrm{m}} = 2.1\ \mathrm{mm}.$$

(*b*) The angular position of the first minimum follows from Eq. 42-1, or

$$\sin \theta = \frac{\lambda}{a} = \frac{480 \times 10^{-9}\,\mathrm{m}}{25 \times 10^{-6}\,\mathrm{m}} = 0.0192.$$

This is so small that, with little error, we can put $\sin \theta \approx \tan \theta$, so

$$\begin{aligned} y &= D \tan \theta \approx D \sin \theta = (52 \times 10^{-2}\,\mathrm{m})(0.0192) \\ &= 10\ \mathrm{mm}. \end{aligned}$$

You can show that there are about 9 fringes in the central peak of the diffraction envelope.

SAMPLE PROBLEM 42-7. What requirements must be met for the central maximum of the envelope of the double-slit interference pattern to contain exactly 11 fringes?

Solution The required condition will be met if the sixth minimum of the interference factor ($\cos^2 \beta$ in Eq. 42-17) coincides with the first minimum of the diffraction factor [$(\sin \alpha/\alpha)^2$ in Eq. 42-17].

The sixth minimum of the interference factor occurs when $d \sin\theta = \frac{11}{2}\lambda$ in Eq. 41-2, or

$$\beta = (11/2)\pi$$

in Eq. 42-14. The first minimum in the diffraction term occurs for $a \sin\theta = \lambda$ (Eq. 42-1), or

$$\alpha = \pi$$

in Eq. 42-16. Dividing Eqs. 42-14 and 42-16 yields $\beta/\alpha = d/a$, and thus

$$\frac{d}{a} = \frac{11}{2}.$$

That is, the slit separation d must be 11/2 times the slit width a. This condition depends only on the ratio of the slit separation d to the slit width a and not at all on the wavelength. For larger λ the pattern is broader than for smaller λ, but there are always 11 fringes in the central peak of the envelope.

Phasor Derivation of Eq. 42-17 (Optional)

Figure 42-19 shows the geometry appropriate for the analysis of the double slit using phasors. Each of the two slits is divided into N zones, as was done for the single slit in Fig. 42-9. The net electric field at P is found from the superposition of the N electric field vectors from the upper slit and the N electric field vectors from the lower slit. The phasor method allows us to combine these contributions to the electric field at P, taking into account their relative phases.

Figure 42-20 shows the first N phasors (corresponding to the upper slit of Fig. 42-19) and their resultant E_1, as in Fig. 42-11. There is a phase difference $\Delta\phi = \phi/N$ between each of the N phasors. To add the second group of N phasors (corresponding to the lower slit) we must find the phase angle ξ between the last phasor from the upper slit and the first phasor from the lower slit. We then draw the N phasors from the lower slit and find their resultant, E_2. The sum of the E_1 and E_2 phasors gives the resultant E_θ that characterizes the double slit.

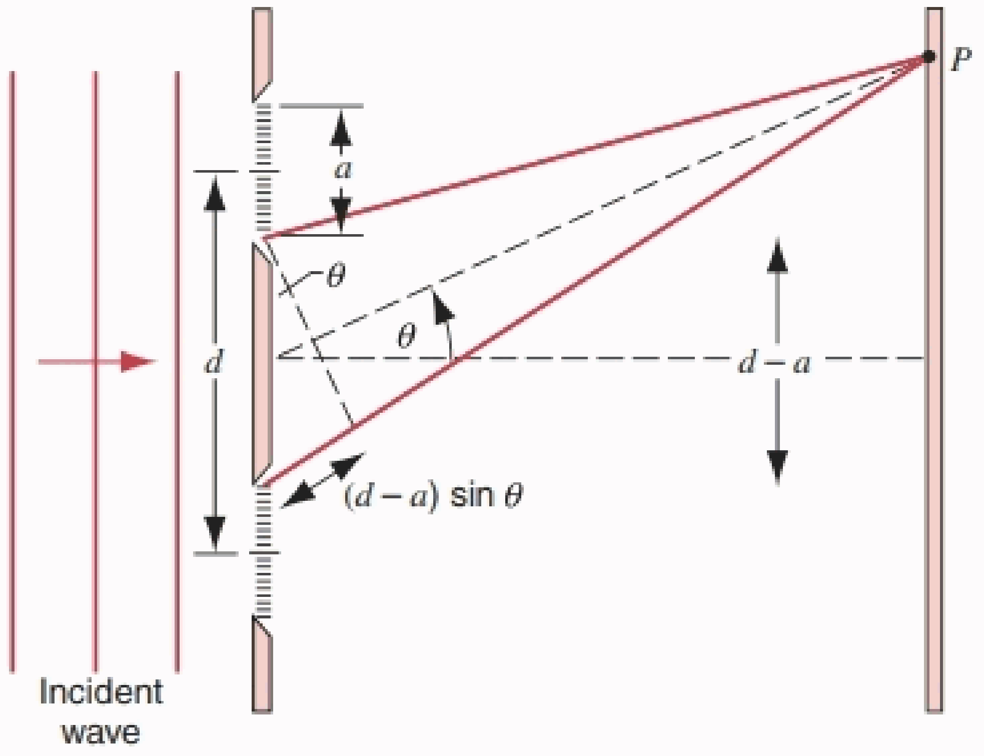


FIGURE 42-19. Each slit in a double slit is divided into N strips. In the differential limit, the strips become infinitesimally small and infinitely numerous. Here, as we did in Fig. 42-9, we show $N = 18$.

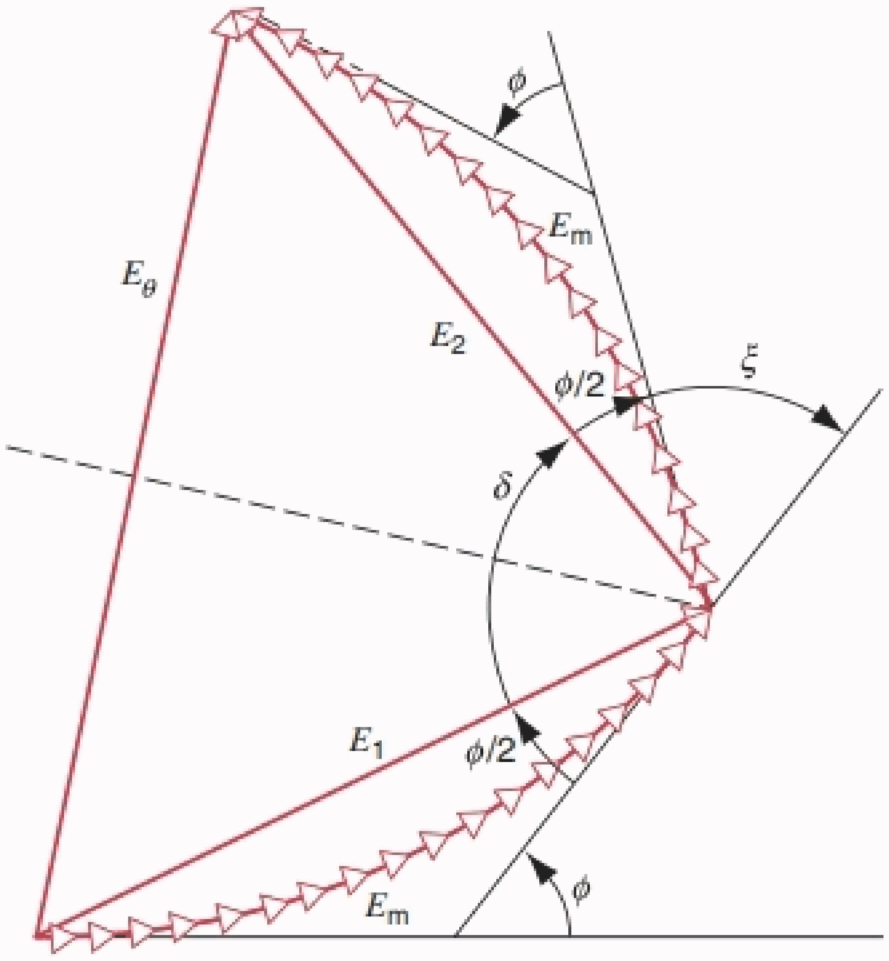


FIGURE 42-20. Phasor diagram used to calculate the resultant electric field in double-slit interference.

From Fig. 42-20, we see that E_θ is the base of an isosceles triangle whose sides have equal lengths E_1 and E_2, which are given by Eq. 42-5. From the geometry of Fig. 42-20,

$$E_\theta = 2E_1 \sin\frac{\delta}{2}, \tag{42-18}$$

where δ, the apex angle of the triangle, can be found from

$$\frac{\phi}{2} + \delta + \frac{\phi}{2} + \xi = \pi, \tag{42-19}$$

which gives

$$\delta = \pi - (\xi + \phi). \tag{42-20}$$

Using Eq. 42-20 to evaluate $\sin \delta/2$, we find

$$\sin\frac{\delta}{2} = \sin\left(\frac{\pi}{2} - \frac{\xi + \phi}{2}\right) = \cos\frac{\xi + \phi}{2}. \tag{42-21}$$

From the expression

$$\frac{\text{path difference}}{\lambda} = \frac{\text{phase difference}}{2\pi}$$

with the phase difference between the two rays (from the bottom of the upper slit and the top of the lower slit, as shown in Fig. 42-19) of ξ and the path difference of $(d - a)\sin\theta$, we have

$$\frac{\xi}{2} = \frac{\pi}{\lambda}(d - a)\sin\theta.$$

Combining this with Eq. 42-7, $\phi/2 = (\pi a/\lambda)\sin\theta$, we find

$$\frac{\xi + \phi}{2} = \frac{\pi}{\lambda}\, d \sin\theta,$$

which is simply β, according to Eq. 42-14. Substituting into Eq. 42-21, we find

$$\sin\frac{\delta}{2} = \cos\beta.$$

Inserting this result into Eq. 42-18 and using Eq. 42-5 for the magnitude of E_1 (or E_2), we obtain

$$E_\theta = 2E_m \frac{\sin\alpha}{\alpha}\cos\beta. \tag{42-22}$$

Squaring Eq. 42-22 gives the intensities as

$$I_\theta = I_m(\cos\beta)^2\left(\frac{\sin\alpha}{\alpha}\right)^2,$$

which is identical to Eq. 42-17. Note that, as was the case in Eq. 41-13, $I_m = 4I_0$.

MULTIPLE CHOICE

42-1 Diffraction and the Wave Theory of Light

1. Which of these statements is most correct?
- (A) Diffraction occurs only for transverse waves.
- (B) Diffraction is proof that light can behave like a wave.
- (C) Diffraction explains rainbows.
- (D) Red sunsets are a diffraction phenomena.

42-2 Single-Slit Diffraction

2. Compare the bright and dark bands from a single slit of width a to the bright and dark bands from a double slit of separation $a/2$. (*a*) Is there a central maximum? (*b*) In the small-angle approximation, are all of the bright and dark bands equally spaced? (*c*) Are both the maxima and minima located at angles given by $\sin\theta \propto k$ where k is an integer?
- (A) Yes, for both
- (B) Only for single-slit diffraction
- (C) Only for double-slit interference
- (D) No, not for either

3. The diffraction pattern of a single slit is observed with green light. The green light is then replaced by red light. Which of the following changes will allow the locations of the diffraction minima on the screen to remain unaffected by the change of light color?
- (A) Increase the width of the slit.
- (B) Decrease the width of the slit.
- (C) Move the lens–screen combination away from the slit.
- (D) Increase the intensity of light source.

4. Two colors of light, red and violet, are allowed to pass through a single slit. It is observed that for a certain slit width, a, the second maximum of the violet light and the first maximum of the red light overlap. The slit width is then decreased. What happens to the overlapping violet and red maxima?
- (A) The maxima separate, with the angle red maximum increasing faster than the violet.
- (B) The maxima separate, with the angle of the violet maximum increasing faster than the red.
- (C) The angle of the red maximum increases while the angle of the violet decreases.
- (D) The angle for both maxima increases so they continue to overlap.

5. The full spectrum of visible light is allowed to pass through a single slit. Where, if any, are regions that are *not* illuminated by some color of visible light?
- (A) Between the central maximum and the first maximum for red
- (B) Between the nth maximum for violet and the $(n + 1)$th maximum for red, where n is any integer
- (C) Between the nth maximum for violet light and the $(n + 1)$th maximum for red, where $n > 3$.
- (D) Every angle corresponds to a maximum for some color

42-3 Intensity in Single-Slit Diffraction

6. The intensity of the central maximum in single slit diffraction
- (A) is independent of the slit width.
- (B) is proportional to the slit width.
- (C) is proportional to the square of the slit width.
- (D) is inversely proportional to the slit width.

7. The intensity of the first maximum beyond the central maximum is approximately what percentage of the intensity of the central maximum?

(A) 50% (B) 25% (C) 5%

(D) It depends on the width of the slit.

8. Diffraction pattern minima are located at angles θ_m, where m is a positive, nonzero integer and

(A) $a\sin\theta_m = m\lambda$. (B) $a\sin\theta_m > m\lambda$.

(C) $a\sin\theta_m < m\lambda$.

9. Diffraction pattern maxima are located at angles θ_m, where m is a nonnegative integer and
- (A) $a\sin\theta_m = (m + 1/2)\lambda$.
- (B) $a\sin\theta_m > (m + 1/2)\lambda$.
- (C) $a\sin\theta_m < (m + 1/2)\lambda$.

42-4 Diffraction at a Circular Aperture

10. Consider diffraction from a square aperture of side d. The first minimum will occur at $d\sin\theta = k\lambda$, where k is a constant that can be found exactly by integration. What value would you expect for k?

(A) $k < 1$ (B) $k = 1$ (C) $1 < k < 1.22$

(D) $k = 1.22$ (E) $1.22 < k$

11. What is the smallest structure that can be resolved with a traditional optical microscope?

(A) A bacterium (B) A virus

(C) A water molecule

(D) There is no limit; someday we will design better microscopes.

42-5 Double-Slit Interference and Diffraction Combined

12. The double-slit pattern of Fig. 42-21*a* seen with a monochromatic light source is somehow changed to the pattern of Fig. 42-21*b*. Consider the following possible changes in conditions:

(A) the wavelength of the light was increased,
(B) the wavelength of the light was decreased,
(C) the width of each slit was increased,
(D) the width of each slit was decreased,
(E) the separation of the slits was increased,
(F) the separation of the slits was decreased.

Which selection(s) could explain the alteration of the pattern?

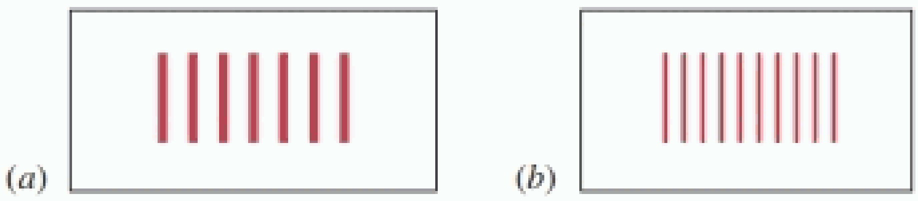


FIGURE 42-21. Multiple-choice question 12.

QUESTIONS

1. Distinguish between Fresnel and Fraunhofer diffraction. Do different physical principles underlie them? If so, what are they? If the same broad principle underlies them, what is it?

2. In what way are interference and diffraction similar? In what way are they different?

3. Suppose that you hold a single, narrow, vertical slit in front of the pupil of your eye and look at a distant light source in the form of a long heated filament. Is the diffraction pattern that you see a Fresnel or a Fraunhofer pattern?

4. Do diffraction effects occur for virtual images as well as for real images? Explain.

5. Do diffraction effects occur for images formed by (*a*) plane mirrors and (*b*) spherical mirrors? Explain.

6. Comment on this statement: "Diffraction occurs in all regions of the electromagnetic spectrum." Consider the x-ray region and the microwave region, for example, and give arguments for believing the statement to be true or false.

7. We have claimed (correctly) that Maxwell's equations predict all the classical optical phenomena. Yet in Chapter 41 (Interference) and in this chapter (Diffraction), there is little mention of Maxwell's equations. Is there an inconsistency here? Where is the impact of Maxwell's equations felt? Discuss.

8. If we were to redo our analysis of the properties of lenses in Chapter 40 by the methods of geometrical optics but without restricting our consideration to paraxial rays and to "thin" lenses, would diffraction phenomena emerge from the analysis? Discuss.

9. Why is the diffraction of sound waves more evident in daily experience than the diffraction of light waves?

10. Sound waves can be diffracted. About what width of a single slit should you use if you wish to broaden the distribution of an incident plane sound wave of frequency 1 kHz?

11. Why do radio waves diffract around buildings, although light waves do not?

12. A loudspeaker horn, used at a rock music concert, has a rectangular aperture 1 m high and 30 cm wide. Will the pattern of sound intensity be broader in the horizontal plane or in the vertical?

13. A particular radar antenna is designed to give accurate measurements of the altitude of an aircraft but less accurate measurements of its direction in a horizontal plane. Must the height-to-width ratio of the radar antenna be less than, equal to, or greater than unity?

14. Describe what happens to a Fraunhofer single-slit diffraction pattern if the whole apparatus is immersed in water.

15. In single-slit diffraction, what is the effect of increasing (*a*) the wavelength and (*b*) the slit width?

16. While listening to the car radio, you may have noticed that the AM signal fades, but the FM signal does not, when you drive under a bridge. Could diffraction have anything to do with this?

17. What will the single-slit diffraction pattern look like if $\lambda > a$?

18. What would the pattern on a screen formed by a double slit look like if the slits did not have the same width? Would the location of the fringes be changed?

19. The shadow of a vertical flagpole cast by the Sun has clearly defined edges near its base, but less well defined edges near its top end. Why?

20. Sunlight falls on a single slit of width 1 pm. Describe qualitatively what the resulting diffraction pattern looks like.

21. In Fig. 42-8, rays r_1 and r_3 are in phase, as are r_2 and r_4. Why is there not a maximum intensity at P_2 rather than a minimum?

22. When we speak of diffraction by a single slit we imply that the width of the slit must be much less than its length. Suppose that, in fact, the length was equal to twice the width. Make a rough guess at what the diffraction pattern would look like.

23. In Fig. 42-6 the optical path lengths from the slit to point P_0 are all the same. Why?

24. In Fig. 42-10*d*, why is E_θ, which represents the first maximum beyond the central maximum, not vertical? (Hint: Consider the effects of a slight winding or unwinding of the coil of phasors in this figure.) See Exercise 12.

25. Give at least two reasons why the usefulness of large telescopes increases as we increase the lens diameter.

26. Are diffraction effects associated with reflecting telescopes, such as the Hubble Space Telescope, that use mirrors instead of lenses? If so, why do we go to the effort of putting such telescopes in space?

27. We have seen that diffraction limits the resolving power of optical telescopes (see Fig. 42-14). Does it also do so for large radio telescopes?

28. Diffraction is more of a nuisance in a telescope than in a camera. Why?

29. In double-slit interference patterns such as that of Fig. 42-16*a*, we said that the interference fringes were modulated in intensity by the diffraction pattern of a single slit. Could we reverse this statement and say that the diffraction pattern of a single slit is intensity modulated by the interference fringes? Discuss.

EXERCISES

42-1 Diffraction and the Wave Theory of Light

42-2 Single-Slit Diffraction

1. When monochromatic light is incident on a slit 0.022 mm wide, the first diffraction minimum is observed at an angle of 1.8° from the direction of the incident beam. Find the wavelength of the incident light.

2. Can you demonstrate the wave nature of x rays by diffracting them through a single slit? Determine the maximum slit width that could be used if a central maximum angular width of 0.12 mrad can just be detected and you guess the wavelength of the x rays to be 0.10 nm.

3. Monochromatic light of wavelength 441 nm falls on a narrow slit. On a screen 2.16 m away, the distance between the second minimum and the central maximum is 1.62 cm. (*a*) Calculate the angle of diffraction θ of the second minimum. (*b*) Find the width of the slit.

4. Light of wavelength 633 nm is incident on a narrow slit. The angle between the first minimum on one side of the central maximum and the first minimum on the other side is 1.97°. Find the width of the slit.

5. A single slit is illuminated by light whose wavelengths are λ_a and λ_b, so chosen that the first diffraction minimum of the λ_a component coincides with the second minimum of the λ_b component. (*a*) What relationship exists between the two wavelengths? (*b*) Do any other minima in the two patterns coincide?

6. A plane wave, with wavelength of 593 nm, falls on a slit of width 420 μm. A thin converging lens, having a focal length of 71.4 cm, is placed behind the slit and focuses the light on a screen. Find the distance on the screen from the center of the pattern to the second minimum.

7. In a single-slit diffraction pattern, the distance between the first minimum on the right and the first minimum on the left is 5.20 mm. The screen on which the pattern is displayed is 82.3 cm from the slit and the wavelength is 546 nm. Calculate the slit width.

8. The distance between the first and fifth minima of a single-slit pattern is 0.350 mm with the screen 41.3 cm away from the slit, using light having a wavelength of 546 nm. (*a*) Calculate the diffraction angle θ of the first minimum. (*b*) Find the width of the slit.

9. A slit 1.16 mm wide is illuminated by light of wavelength 589 nm. The diffraction pattern is seen on a screen 2.94 m away. Find the distance between the first two diffraction minima on the same side of the central maximum.

42-3 Intensity in Single-Slit Diffraction

10. If you double the width of a single slit, the intensity of the central maximum of the diffraction pattern increases by a factor of four, even though the energy passing through the slit only doubles. Explain this quantitatively.

11. Monochromatic light with wavelength 538 nm falls on a slit with width 25.2 μm. The distance from the slit to a screen is 3.48 m. Consider a point on the screen 1.13 cm from the central maximum. (*a*) Calculate θ. (*b*) Calculate α. (*c*) Calculate the ratio of the intensity at this point to the intensity at the central maximum.

12. In Fig. 42-10*d*, calculate the angle E_θ makes with the vertical; see Question 24 and Problem 3.

13. Calculate the width of the central maximum in a single-slit diffraction pattern in which $a = 10\lambda$. Compare your result with Fig. 42-12*c*. See Sample Problem 42-4.

42-4 Diffraction at a Circular Aperture

14. An astronaut in a satellite claims to be able to just barely resolve two point sources on Earth, 163 km below. Calculate the (*a*) angular and (*b*) linear separation of the two point sources, assuming ideal conditions. Take $\lambda = 540$ nm and the pupil diameter of the astronaut's eye to be 4.90 mm.

15. The two headlights of an approaching automobile are 1.42 m apart. At what (*a*) angular separation and (*b*) maximum distance will the eye resolve them? Assume a pupil diameter of 5.00 mm and a wavelength of 562 nm. Also assume that diffraction effects alone limit the resolution.

16. The wall of a large room is covered with acoustic tile in which small holes are drilled 5.20 mm from center to center. How far can a person be from such a tile and still distinguish the individual holes, assuming ideal conditions? Assume the diameter of the pupil of the observer's eye to be 4.60 mm and the wavelength to be 542 nm.

17. Find the separation of two points on the Moon's surface that can just be resolved by the 200-in. (5.08-m) telescope at Mount Palomar, assuming that this distance is determined by diffraction effects. Assume a wavelength of 565 nm.

18. A navy cruiser employs radar with a wavelength of 1.57 cm. The circular antenna has a diameter of 2.33 m. At a range of 6.25 km, what is the smallest distance that two speedboats can be from each other and still be resolved as two separate objects by the radar system?

19. If Superman really had x-ray vision at 0.12-nm wavelength and a 4.3-mm pupil diameter, at what maximum altitude could he distinguish villains from heroes assuming the minimum detail required was 4.8 cm?

20. A "spy in the sky" satellite orbiting at 160 km above the Earth's surface has a lens with a focal length of 3.6 m. Its resolving power for objects on the ground is 30 cm; it could easily measure the size of an aircraft's air intake. What is the effective lens diameter, determined by diffraction consideration alone? Assume $\lambda = 550$ nm. Far more effective satellites are reported to be in operation today.

21. The paintings of Georges Seurat consist of closely spaced small dots (≈ 2 mm in diameter) of pure pigment, as indicated in Fig. 42-22. The illusion of color mixing occurs because the pupils of the observer's eyes diffract light entering them. Calculate the minimum distance an observer must stand from such a painting to achieve the desired blending of color. Take the wavelength of the light to be 475 nm and the diameter of the pupil to be 4.4 mm.

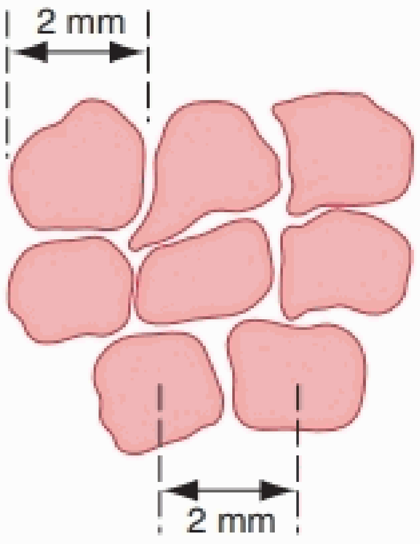


FIGURE 42-22. Exercise 21.

22. In June 1985, a laser beam was fired from the Air Force Optical Station on Maui, Hawaii, and reflected back from the shuttle *Discovery* as it sped by, 354 km overhead. The diameter of the central maximum of the beam at the shuttle position was said to be 9.14 m and the beam wavelength was 500 nm. What is the effective diameter of the laser aperture at the Maui ground station? (Hint: A laser beam spreads because of diffraction; assume a circular exit aperture.)

23. (*a*) A circular diaphragm 60 cm in diameter oscillates at a frequency of 25 kHz in an underwater source of sound used for submarine detection. Far from the source, the sound intensity is distributed as a diffraction pattern for a circular hole whose diameter equals that of the diaphragm. Take the speed of sound in water to be 1450 m/s and find the angle between the normal to the diaphragm and the direction of the first minimum. (*b*) Repeat for a source having an (audible) frequency of 1.0 kHz.

24. Millimeter-wave radar generates a narrower beam than conventional microwave radar. This makes it less vulnerable to antiradar missiles. (*a*) Calculate the angular width, from first minimum to first minimum, of the central "lobe" produced by a 220-GHz radar beam emitted by a 55-cm-diameter circular antenna. (The frequency is chosen to coincide with a low-absorption atmospheric "window.") (*b*) Calculate the same quantity for the ship's radar described in Exercise 18.

42-5 Double-Slit Interference and Diffraction Combined

25. Two slits of width a and separation d are illuminated by a coherent beam of light of wavelength λ. What is the linear separation of the interference fringes observed on a screen that is a distance D away?

26. (*a*) Design a double-slit system in which the fourth fringe, not counting the central maximum, is missing. (*b*) What other fringes, if any, are also missing?

27. (*a*) For $d = 2a$ in Fig. 42-23, how many interference fringes lie in the central diffraction envelope? (*b*) If we put $d = a$, the two slits coalesce into a single slit of width $2a$. Show that Eq. 42-17 reduces to the diffraction pattern for such a slit.

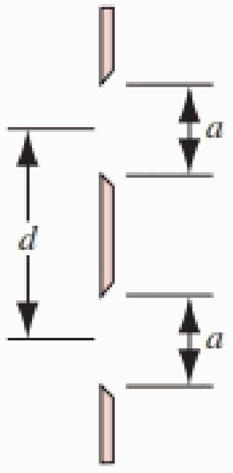


FIGURE 42-23. Exercise 27.

28. Suppose that, as in Sample Problem 42-7, the envelope of the central peak contains 11 fringes. How many fringes lie between the first and second minima of the envelope?

29. (*a*) How many complete fringes appear between the first minima of the fringe envelope on either side of the central maximum for a double-slit pattern if $\lambda = 557$ nm, $d = 0.150$ mm, and $a = 0.030$ mm? (*b*) What is the ratio of the intensity of the third fringe to the side of the center to that of the central fringe?

PROBLEMS

1. Manufacturers of wire (and other objects of small dimensions) sometimes use a laser to continually monitor the thickness of the product. The wire intercepts the laser beam, producing a diffraction pattern like that of a single slit of the same width as the wire diameter; see Fig. 42-24. Suppose a He-Ne laser, wavelength 632.8 nm, illuminates a wire, the diffraction pattern being projected onto a screen 2.65 m away. If the desired wire diameter is 1.37 mm, what would be the observed distance between the two tenth-order minima on each side of the central maximum?

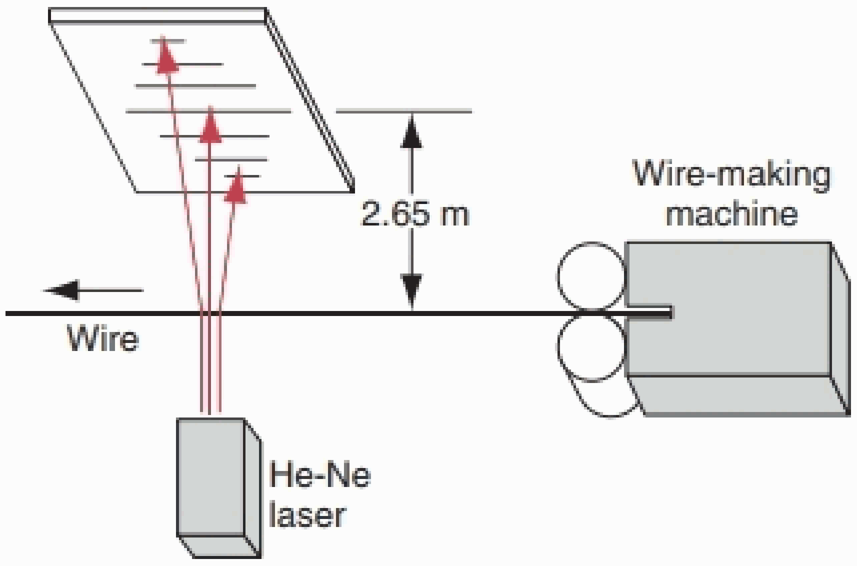


FIGURE 42-24. Problem 1.

2. A monochromatic beam of parallel light is incident on a "collimating" hole of diameter $a \gg \lambda$. Point P lies in the geometrical shadow region on a distant screen, as shown in Fig. 42-25*a*. Two obstacles, shown in Fig. 42-25*b* are placed in turn over the collimating hole. A is an opaque circle with a hole in it and B is the "photographic negative" of A. Using superposition concepts, show that the intensity at P is identical for each of the two diffracting objects A and B (this is called *Babinet's principle*). In this connection, it can be shown that the diffraction pattern of a wire is the same as that of a slit of equal width. See "Measuring the Diameter of a Hair by Diffraction," by S. M. Curry and A. L. Schawlow, *American Journal of Physics,* May 1974, p. 412.

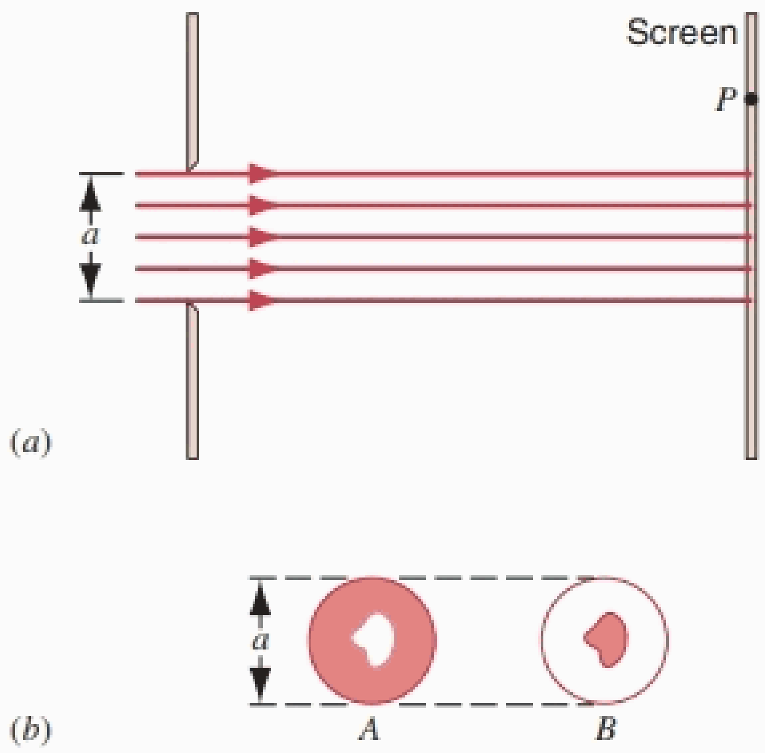


FIGURE 42-25. Problem 2.

3. (*a*) Show that the values of α at which intensity maxima for single-slit diffraction occur can be found exactly by differentiating Eq. 42-8 with respect to α and equating to zero, obtaining the condition $\tan \alpha = \alpha$. (*b*) Find the values of α satisfying this relation by plotting graphically the curve $y = \tan \alpha$ and the straight line $y = \alpha$ and finding their intersections or by using a pocket calculator to find an appropriate value of α by trial and error. (*c*) Find the (nonintegral) values of m corresponding to successive maxima in the single-slit pattern. Note that the secondary maxima do not lie exactly halfway between minima.

4. In a Soviet–French experiment to monitor the Moon's surface with a light beam, pulsed radiation from a ruby laser ($\lambda = 0.69$ μm) was directed to the Moon through a reflecting telescope with a mirror radius of 1.3 m. A reflector on the Moon behaved like a circular plane mirror with radius 10 cm, reflecting the light directly back toward the telescope on Earth. The reflected light was then detected by a photometer after being brought to a focus by this telescope. What fraction of the original light energy was picked up by the detector? Assume that for each direction of travel all the energy is in the central diffraction circle.

5. It can be shown that, except for $\theta = 0$, a circular obstacle produces the same diffraction pattern as a circular hole of the same diameter. Furthermore, if there are many such obstacles, such as water droplets located randomly, then the interference effects vanish, leaving only the diffraction associated with a single obstacle. (*a*) Explain why one sees a "ring" around the Moon on a foggy night. The ring is usually reddish in color. Explain why. (*b*) Calculate the size of the water droplets in the air if the ring around the Moon appears to have a diameter 1.5 times that of the Moon. The angular diameter of the Moon in the sky is 0.5°. (*c*) At what distance from the Moon might a bluish ring be seen? Sometimes the rings are white. Why? (*d*) The color arrangement is opposite to that in a rainbow. Why should this be so?

6. An acoustic double-slit system (slit separation d, slit width a) is driven by two loudspeakers as shown in Fig. 42-26. By use of a variable delay line, the phase of one of the speakers may be varied. Describe in detail what changes occur in the intensity pattern at large distances as this phase difference is varied from zero to 2π. Take both interference and diffraction effects into account.

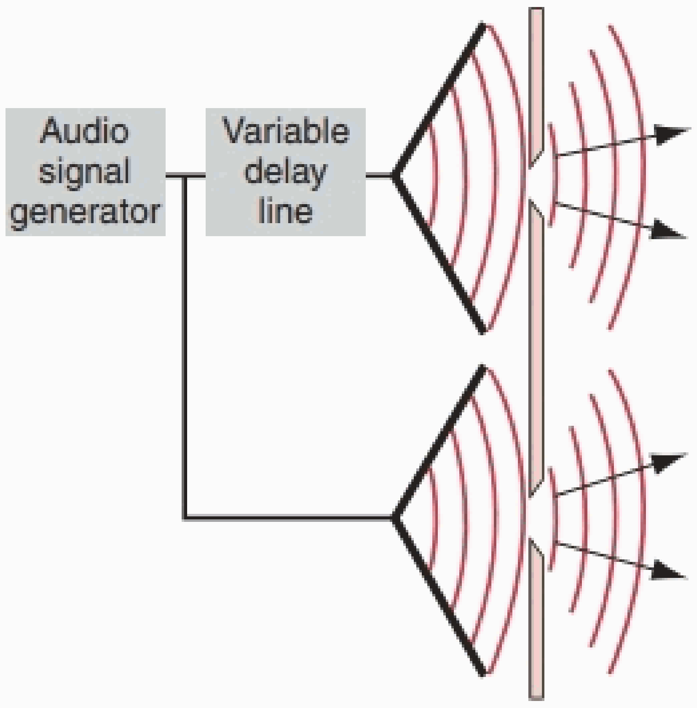


FIGURE 42-26. Problem 6.

7. Light of wavelength 440 nm passes through a double slit, yielding the diffraction pattern of intensity I versus deflection angle θ shown in Fig. 42-27. Calculate (*a*) the slit width and (*b*) the slit separation. (*c*) Verify the displayed intensities of the $m = 1$ and $m = 2$ interference fringes.

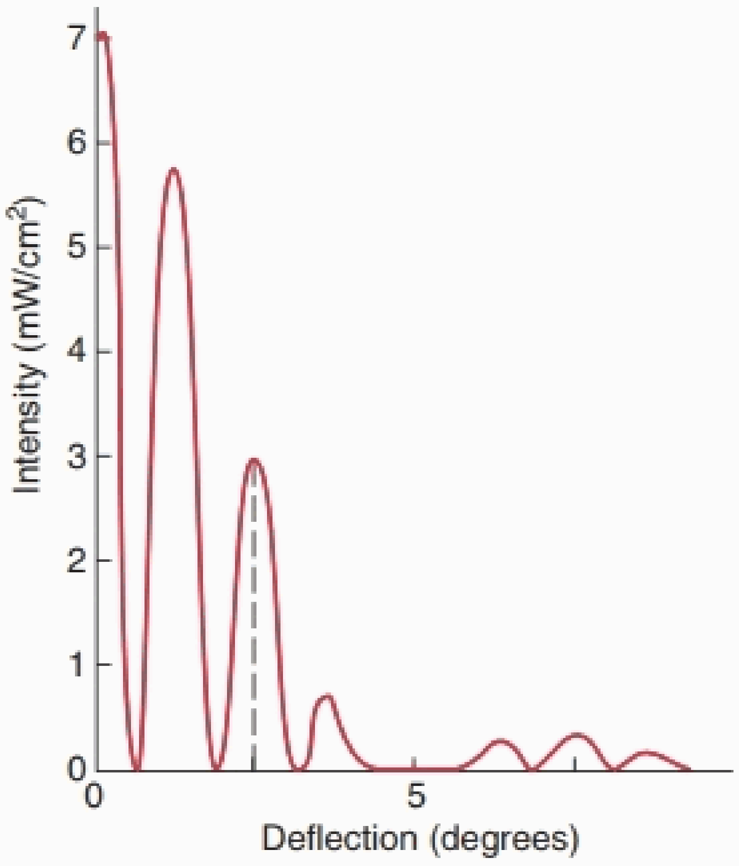


FIGURE 42-27. Problem 7.

COMPUTER PROBLEMS

1. Numerically generate a graph of intensity as a function of angle from the central axis for diffraction from a circular aperture. Assume an aperture diameter $d = 5\lambda$. Do this by assuming that the circular aperture can be approximated by a square lattice of point sources, each separated by a distance of $d/10$. Of course, only lattice points within the circle are allowed to radiate! Verify the coefficient 1.22 in Eq. 42-11.

2. Numerically integrate Eq. 42-8 for various values of a and λ to verify that the total power incident on the single slit is the same as the power incident on the screen. Some values to try: $a/\lambda = 0.5$, 1, and 2.

CHAPTER 43

GRATINGS AND SPECTRA

In Chapter 41 we discussed the interference pattern produced when monochromatic light is incident on a double slit: a pattern of bright and dark bands (interference fringes) is produced. Each of the slits is regarded as a source of diffracted waves. In this chapter we extend our discussion to cases in which the number of diffraction sources is larger (often much larger) than two. We consider multiple arrays of slits in a plane and also three-dimensional arrays of atoms in a solid (for which we use x rays rather than visible light).

43-1 MULTIPLE SLITS

Just as Thomas Young used his double-slit interference technique for the first measurement of the wavelength of light, we continue today to use a variety of interference techniques for precise measurement of wavelengths. For example, from 1960 to 1983, the standard meter was defined in terms of the wavelength of the light emitted by atoms of krypton (see Section 1-4). To enable its use as a standard, it was necessary to determine the value of the wavelength to nine significant figures!

In principle, we can use a double-slit interference pattern to measure the wavelength. If we know the slit separation d, then by locating the bright fringes on a screen and using Eq. 41-1 ($d \sin \theta = m\lambda$) we can find the wavelength. Often it is more convenient to determine the separation between adjacent maxima, in which case we can use Eq. 41-4 ($\Delta y = \lambda D/d$) to find λ. The precision of this method is limited by our ability to make an accurate determination of the location or spacing of the fringes. Figure 43-1*a* shows a typical double-slit interference pattern; you can see that trying to measure from the midpoint of one fringe to the midpoint of the next one involves an uncertainty in locating the midpoints.

The interference pattern in Fig. 43-1*b*, which was obtained using five slits, offers a slight improvement; note that the bright fringes are narrower, meaning that we will be able to do a slightly better job of measuring the fringe spacing and thus determine the wavelength more precisely. The separation d between adjacent slits has the same value for both cases, and you can see by comparing the two interference patterns that the spacing of the fringes is determined

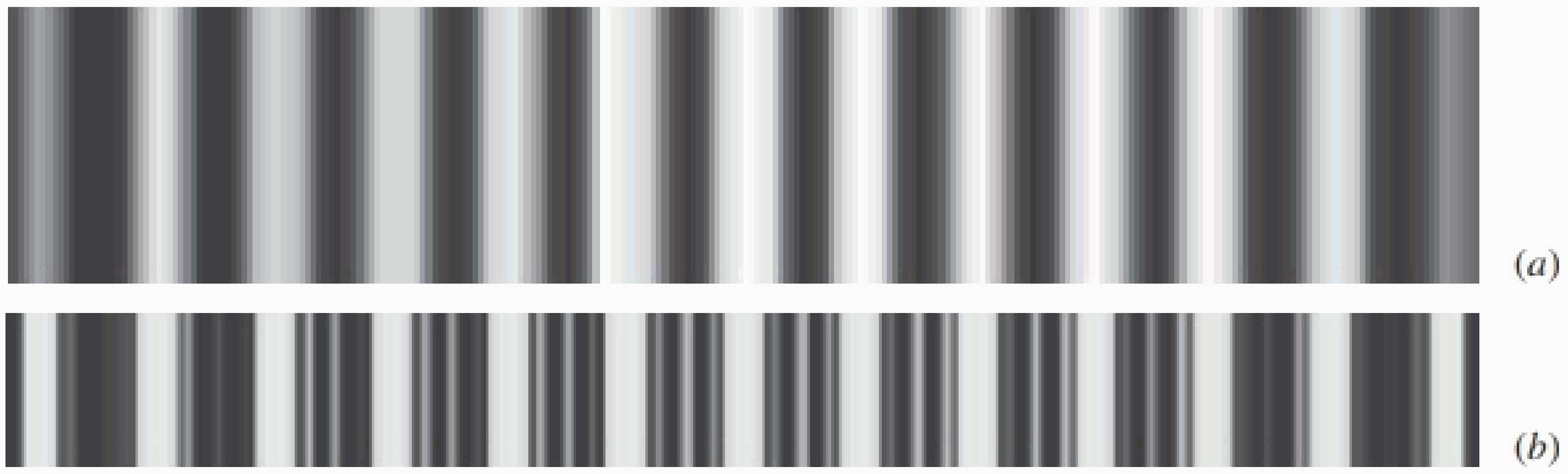


FIGURE 43-1. The diffraction pattern for a grating with (*a*) two slits and (*b*) five slits. Note that, in the case of the five-slit grating, the fringes are sharper (narrower), and secondary maxima of low intensity appear between the bright principal maxima.

by the slit separation d, independently of the number of slits. As we increase the number of slits, the bright fringes continue to become narrower, and the precision of our wavelength measurement continues to improve. Such a multiple-slit arrangement, in which the number of slits N may be as large as 10^4, is called a *diffraction grating.*

From Fig. 43-1*b* you can see a second effect of increasing the number of slits—faint *secondary maxima* (three in Fig. 43-1*b*) appear between adjacent bright fringes. As the number of slits increases, the number of secondary maxima increases but their brightness decreases; when N is large we can neglect the secondary maxima completely.

In this chapter we consider the interference patterns for diffraction gratings. We will consider how the sharpness of the fringes depends on the number of slits in the grating, and we will explain the origin of the secondary maxima. As we did in Chapter 41, we consider only the Fraunhofer arrangement, in which there is an assumed infinite distance between the light source and the slits as well as between the slits and the screen. Equivalently, parallel light is incident on the slits, and parallel light emerges from the slits (perhaps to be focused by a lens) to form an image on the screen.

Figure 43-2 shows a five-slit grating (compare Fig. 41-6). A principal maximum occurs when the path difference between rays from any pair of adjacent slits, which is given by $d \sin \theta$, is equal to an integral number of wavelengths, or

$$d \sin \theta = m\lambda \qquad m = 0, \pm 1, \pm 2, \ldots, \tag{43-1}$$

where m is called the *order number.* Equation 43-1 is identical with Eq. 41-1 for the maxima of the double slit. Note that if light passing through any pair of *adjacent* slits is in phase at a particular point on the screen, then light passing through any pair of slits, even nonadjacent ones, is also in phase at that point. For a given slit separation d, the locations of the principal maxima are determined by the wavelength, according to Eq. 43-1. The locations of the principal maxima are independent of the number of slits N, which, as we shall see, determines the width or sharpness of the principal maxima. For slits of width a, the relative intensities of the principal maxima within the diffraction envelope are determined by the ratio a/λ, which does not affect their locations. A comparison of the intensity patterns for two-slit and five-slit gratings is given in Fig. 43-3.

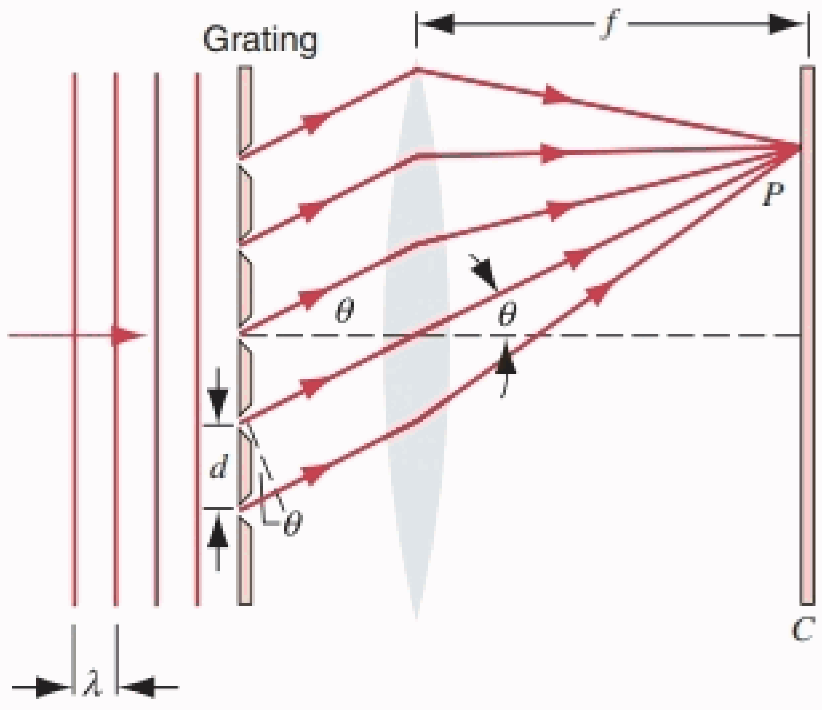


FIGURE 43-2. An idealized diffraction grating containing five slits. The slit width a is assumed to be much smaller than λ, although this condition may not be realized in practice. Also, the focal length f will in practice be much greater than d; the figure distorts these dimensions for clarity.

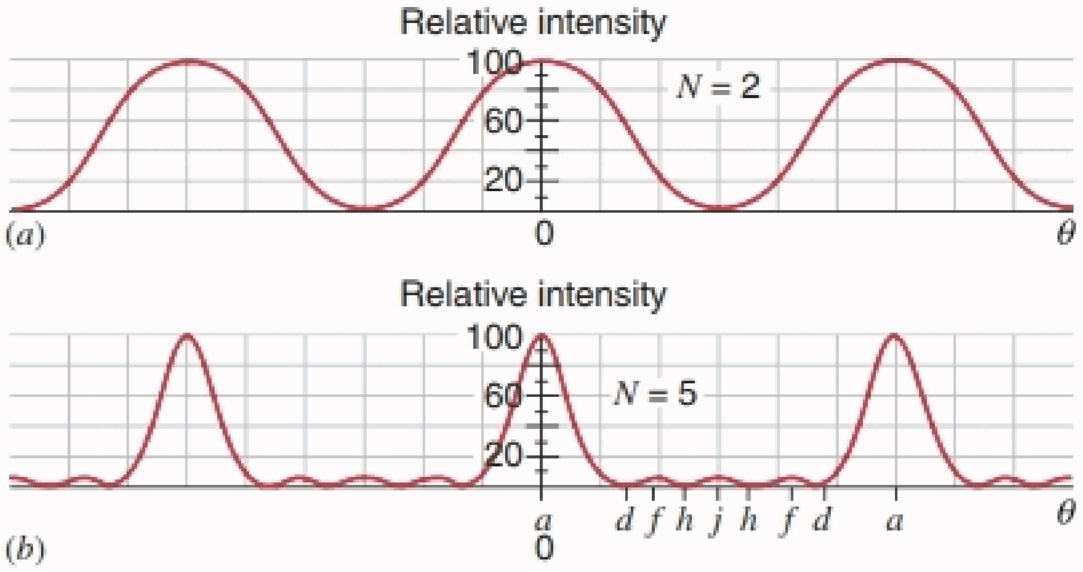


FIGURE 43-3. Calculated intensity patterns for (*a*) a two-slit and (*b*) a five-slit grating, having the same values of d and λ. Note the sharpening of the principal maxima and the appearance of faint secondary maxima in (*b*); compare with Fig. 43-1. The letters in (*b*) refer to Fig. 43-6. This calculation does not include diffraction effects due to the slit width; that is, we assume we are near the central region of Fig. 43-1 where the principal maxima have essentially equal intensities.

Width of the Maxima

The sharpening of the principal maxima as N is increased can be understood by a graphical argument, using phasors. Figures 43-4*a* and 43-4*b* show conditions at the central principal maximum for a two-slit and a five-slit grating. The small arrows represent the amplitudes of the wave disturbances from each slit arriving at the screen at the position of the central maximum, for which $m = 0$, and thus $\theta = 0$, in Eq. 43-1.

On either side of the central maximum there is a minimum of zero intensity, which lies at an angle $\delta\theta_0$ off the central axis, as shown in Fig. 43-5. Figures 43-4*c* and 43-4*d* show the phasors at this point. The phase difference $\Delta\phi$ between waves from adjacent slits, which is zero at the central principal maximum, must be chosen so that the array of phasors just closes on itself, yielding zero resultant intensity. For $N = 2$, $\Delta\phi = 2\pi/2$ (= 180°); for $N = 5$, $\Delta\phi = 2\pi/5$ (= 72°). In the general case it is given by

$$\Delta\phi = \frac{2\pi}{N}. \tag{43-2}$$

This phase difference for adjacent waves corresponds to a path difference ΔL given by

$$\frac{\text{phase difference}}{2\pi} = \frac{\text{path difference}}{\lambda},$$

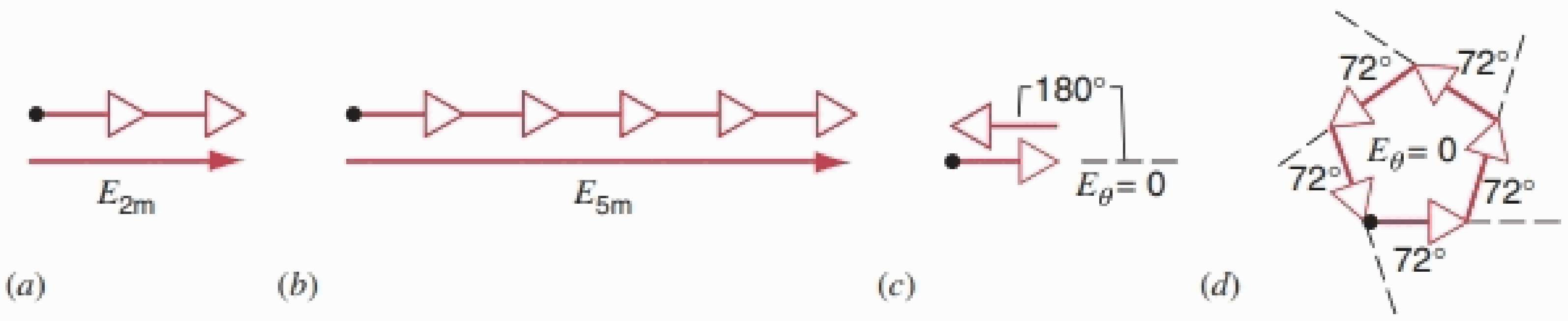


FIGURE 43-4. (*a, b*) The conditions at the central maximum for a two-slit and a five-slit grating, respectively. (*c, d*) The corresponding conditions at a minimum of zero intensity that lies on either side of this central maximum.

or

$$\Delta L = \left(\frac{\lambda}{2\pi}\right)\Delta\phi = \left(\frac{\lambda}{2\pi}\right)\left(\frac{2\pi}{N}\right) = \frac{\lambda}{N}. \quad (43\text{-}3)$$

However, the path difference ΔL at the first minimum (see Figs. 43-2 and 43-5) is also given by $d \sin \delta\theta_0$, so that we can write

$$d \sin \delta\theta_0 = \frac{\lambda}{N},$$

or

$$\sin \delta\theta_0 = \frac{\lambda}{Nd}. \quad (43\text{-}4)$$

Since $N \gg 1$ for actual gratings, $\sin \delta\theta_0$ is ordinarily quite small (that is, the lines are sharp), and to a good approximation we may replace $\sin \delta\theta_0$ by $\delta\theta_0$, expressed in radians, or

$$\delta\theta_0 = \frac{\lambda}{Nd}. \quad (43\text{-}5)$$

This equation shows specifically that if we increase N for a given λ and d, then $\delta\theta_0$ decreases, which means that the central principal maximum becomes sharper.

To obtain the result for *any* principal maximum, we consider the geometry of Fig. 43-5, in which the mth principal maximum occurs at an angle θ. We move away from the maximum through an angular displacement $\delta\theta$ to arrive at the next minimum; we take this angle $\delta\theta$ to be a measure of the angular width of the maximum. At the maximum, the path difference between rays from adjacent slits is $m\lambda$ (see Eq. 43-1). At the next minimum, the path difference between rays from adjacent slits is $m\lambda + \lambda/N$, the additional path length of λ/N being given by Eq. 43-3. For example, consider the case of $N = 10$. The additional path length between adjacent slits at the minimum is 0.1λ. The path difference between slits 1 and 6 is therefore $5(m\lambda + 0.1\lambda) = 5m\lambda + 0.5\lambda$; the path lengths differ by a half-integral number of wavelengths, so the rays interfere destructively. The same is true for slits 2 and 7, slits 3 and 8, and so forth. If the additional path difference is λ/N, then rays from the lower $N/2$ slits undergo pairwise destructive interference with rays from the upper $N/2$ slits.

At the angle $\theta + \delta\theta$, the path difference between rays from adjacent slits is

$$\begin{aligned} d \sin(\theta + \delta\theta) &= d(\sin\theta \cos\delta\theta + \cos\theta \sin\delta\theta) \\ &\approx d \sin\theta + (d\cos\theta)\delta\theta, \end{aligned}$$

where we assume $\delta\theta$ is small, which allows us to approximate $\cos\delta\theta \approx 1$ and $\sin\delta\theta \approx \delta\theta$. Setting this path difference equal to $m\lambda + \lambda/N$, its value at the minimum, we obtain

$$d \sin\theta + (d\cos\theta)\delta\theta = m\lambda + \frac{\lambda}{N}$$

or, using Eq. 43-1,

$$(d\cos\theta)\delta\theta = \frac{\lambda}{N}.$$

Solving for $\delta\theta$ gives

$$\delta\theta = \frac{\lambda}{Nd\cos\theta}. \quad (43\text{-}6)$$

This result gives the angular width* for the principal maximum that occurs at the angle θ, corresponding to the particular order m. Note that Eq. 43-6 reduces to Eq. 43-5 for the central maximum ($\theta = 0$). For a given N, d, and λ, the central maximum is the narrowest ($\cos\theta = 1$); the widths increase as we go to larger θ (and therefore to larger orders m). Equation 43-6 shows that $\delta\theta$ becomes smaller (the maxima become sharper) as the product Nd increases. This product (the number of slits times the distance between slits) gives the total width of the grating. Thus the peaks become sharper as the width of the grating increases.

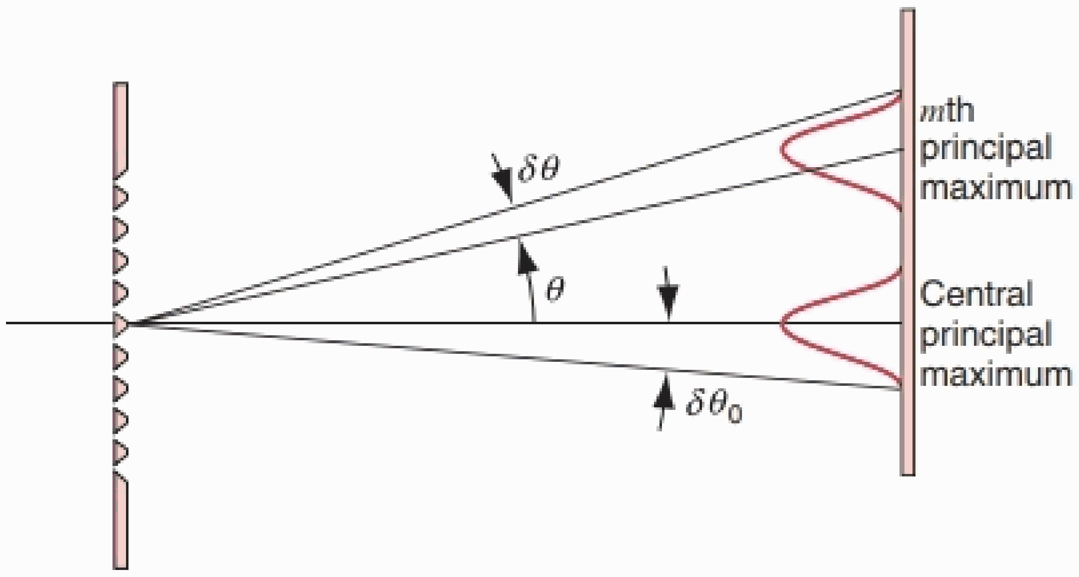


FIGURE 43-5. A principal maximum lies at the position given by the angle θ, and the first minimum occurs at the angle $\delta\theta$ from that maximum. The angle $\delta\theta$ can be taken as a measure of the width or sharpness of the maximum. The width of the central maximum is given by the angle $\delta\theta_0$.

---

* As defined by Eq. 43-6, the width is the angular interval from the peak to the first minimum. The usual definition of the width of a peak is the full interval covered by the peak at half its maximum height (see, for example, Fig. 42-12). These two measures of the width are roughly equal, and we take Eq. 43-6 to represent a measure of the width of the peak.

The Secondary Maxima (Optional)

The origin of the secondary maxima that appear for $N > 2$ can also be understood using the phasor method. Figure 43-6*a* shows conditions for the central principal maximum for a five-slit grating. The phasors are in phase. As we depart from the central maximum, θ in Fig. 43-2 increases from zero and the angle between adjacent phasors increases from zero to $\Delta\phi = (2\pi/\lambda)(d \sin \theta)$. Successive figures show how the resultant wave amplitude E_θ varies with $\Delta\phi$. Verify by graphical construction that a given figure represents conditions for both $\Delta\phi$ and $2\pi - \Delta\phi$. Thus we start at $\Delta\phi = 0$, proceed to $\Delta\phi = 180°$, and then trace backward through the sequence, following the phase differences shown in parentheses, until we reach $\Delta\phi = 360°$. This sequence corresponds to traversing the intensity pattern from the central principal maximum to an adjacent one. Figure 43-6, which should be compared with Fig. 43-3*b*, shows that for $N = 5$ there are three secondary maxima, corresponding to $\Delta\phi = 110°$, 180°, and 250°. Make a similar analysis for $N = 3$ and show that only one secondary maximum occurs. In general, for a grating with N slits, there are $N - 2$ secondary maxima. In actual gratings, which commonly contain 10,000 to 50,000 "slits," the secondary maxima lie so close to the principal maxima or are so reduced in intensity that they cannot be observed experimentally. ■

SAMPLE PROBLEM 43-1. A certain grating has 10^4 slits with a spacing of $d = 2.1\ \mu\text{m} = 2100$ nm. It is illuminated with yellow sodium light ($\lambda = 589$ nm). Find (*a*) the angular position of all principal maxima observed and (*b*) the angular width of the largest-order maximum.

Solution (*a*) From Eq. 43-1, we have

$$\sin\theta = \frac{m\lambda}{d} = \frac{m(589\ \text{nm})}{2100\ \text{nm}},$$

which gives

$$\theta = 16.3°\ (m = 1),\ 34.1°\ (m = 2),\ \text{and}\ 57.3°\ (m = 3),$$

with corresponding values at $\theta < 0$ for $m < 0$. For $m = 4$, $\sin\theta > 1$. Thus $m = 3$ is the highest order observed, which gives a total of seven principal maxima (a central maximum and three on each side of center).

(*b*) For the $m = 3$ maximum, Eq. 43-6 gives

$$\delta\theta = \frac{\lambda}{Nd\cos\theta} = \frac{589\ \text{nm}}{(10^4)(2100\ \text{nm})(\cos 57.3°)}$$
$$= 5.2 \times 10^{-5}\ \text{rad} = 0.0030°.$$

This is an exceedingly narrow principal maximum.

Note that Eq. 43-6, being a dimensionless ratio, gives its result in radian measure. This occurs because we derived Eq. 43-6 using the approximation $\sin\delta\theta \approx \delta\theta$, which is valid only in radian measure.

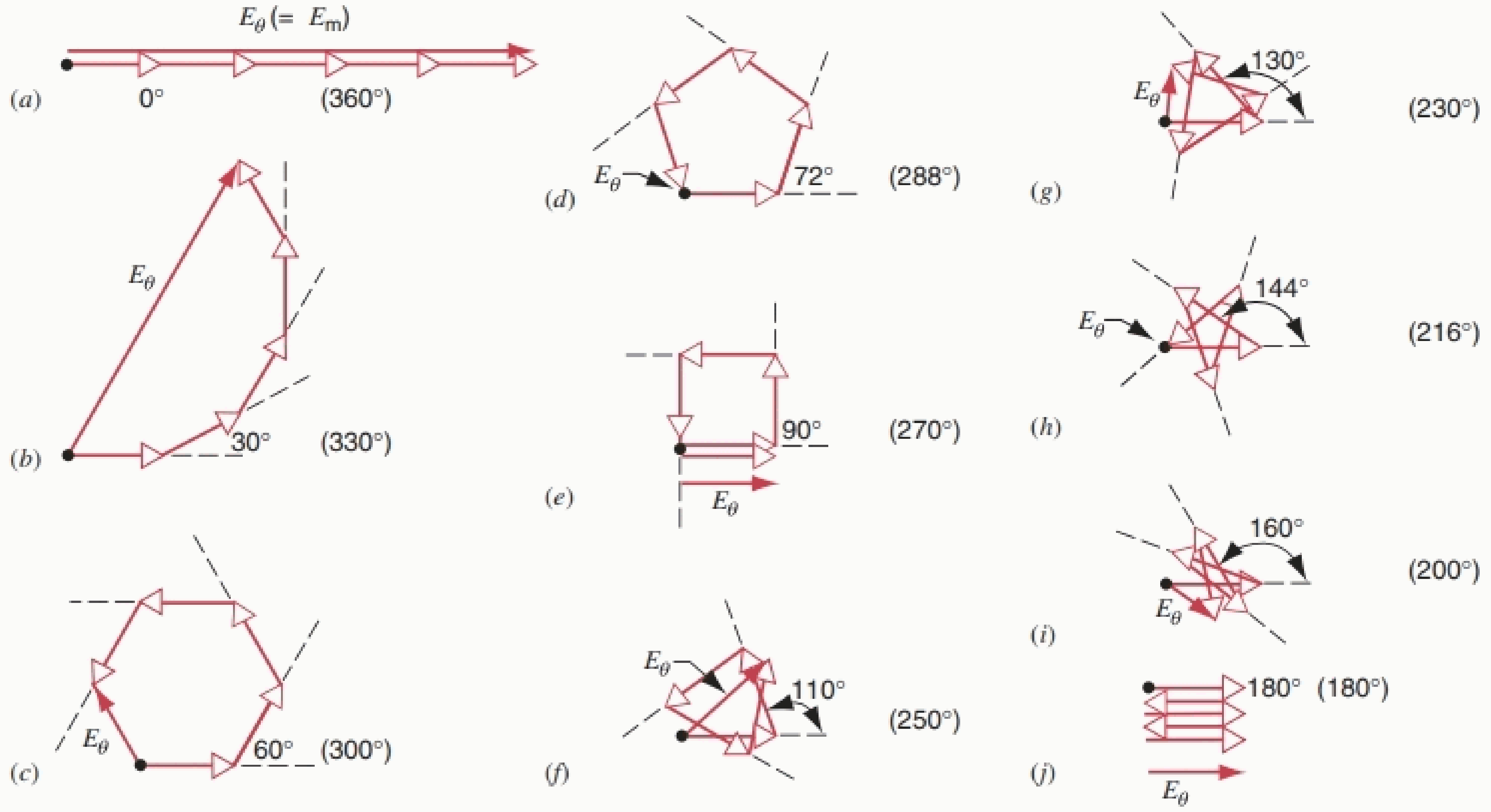


FIGURE 43-6. The figures taken in sequence from (*a*) to (*j*) and then from (*j*) to (*a*) show conditions as the intensity pattern of a five-slit grating is traversed from the central principal maximum to an adjacent principal maximum. Phase differences between waves from adjacent slits are shown directly; those in going from (*j*) to (*a*) are in parentheses. Principal maxima occur at (*a*), secondary maxima at or near (*f*) and (*j*), and minima of zero intensity at (*d*) and (*h*). Compare with Fig. 43-3*b*.

43-2 DIFFRACTION GRATINGS

A typical grating might contain $N = 10{,}000$ slits distributed over a width of a few centimeters, equivalent to a grating spacing d of a few micrometers. As we have seen in Sample Problem 43-1, when Nd is a few centimeters, the maxima are very narrow, which allows their position to be measured with great precision. Gratings are therefore used to determine wavelengths and to study the structure and intensity of the principal maxima.

Any regular periodic structure can serve as a diffraction grating—for example, the grooves of a compact disk, which produce a rainbow pattern when light is reflected from the surface of the disk. Gratings can produce their images by transmitted light, as in Fig. 43-2; there are also *reflection gratings,* which produce their images in reflected light. In the grating of Fig. 43-2, there is a periodic change in the *amplitude* (and no change in phase) of the light as a function of position across the grating. It is also possible to make gratings (of either the reflection or transmission type) that cause a periodic change in the *phase* (and a negligible change in the amplitude) of the light as a function of position across the grating. Most gratings used for visible light, whether of the reflection or transmission type, are phase gratings.

Gratings are made by ruling equally spaced parallel grooves in a thin layer of aluminum or gold deposited on a glass plate, using a diamond cutting point whose motion is automatically controlled by a ruling engine. Once such a master grating has been prepared, replicas can be formed by pouring a liquid plastic on the grating, allowing it to harden, and stripping it off. The stripped plastic, fastened to a flat piece of glass or other backing, forms a good grating.

Figure 43-7 shows a cross section of a common type of reflection phase grating. (If the grating were transparent, it could function as a transmission phase grating, since light passing through different thicknesses will have varying changes in phase.) The angles of the grooves are chosen so that light of a particular order is reflected in a particular direction. In this way the intensity of one particular order can be enhanced over that of other orders. Cutting gratings in this way is called *blazing.* Most gratings in use today are blazed gratings.

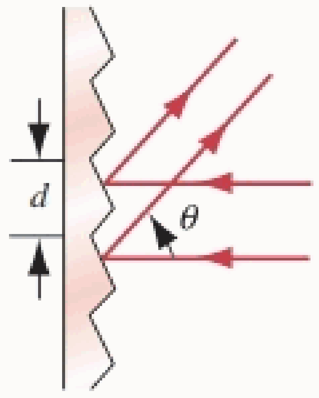


FIGURE 43-7. A cross section of a blazed grating viewed in reflected light. There is a path difference of $d \sin\theta$ between the two rays shown.

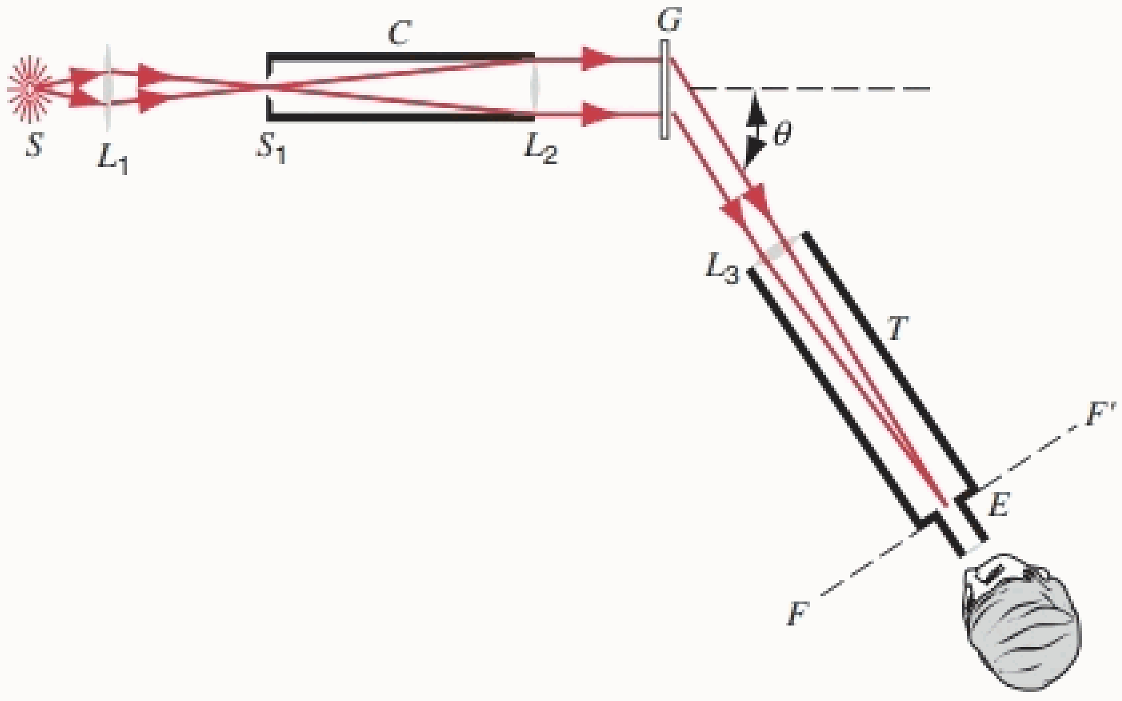


FIGURE 43-8. A simple type of grating spectroscope used to analyze the wavelengths of the light emitted by the source S.

Figure 43-8 shows a simple grating spectroscope, used for viewing the spectrum of a light source, assumed to emit a number of discrete wavelengths. The light from source S is focused by lens L_1 on a slit S_1 placed in the focal plane of lens L_2. The parallel light emerging from collimator C falls on grating G. Parallel rays associated with a particular interference maximum occurring at angle θ fall on lens L_3, and are brought to a focus in plane FF'. The image formed in this plane is examined using a magnifying lens arrangement E (the eyepiece). The entire spectrum can be viewed by rotating telescope T through various angles. Instruments used for scientific research or in industry are more complex than the simple arrangement of Fig. 43-8. They invariably employ photographic or photoelectric recording and are called *spectrographs.* Figure 43-9 shows examples of spectra of visible light recorded by a spectrograph. Each line in the figure is in effect an image of the slit S_1 corresponding to one of the many individual wavelengths emitted from the source. For this reason, such images are called *spectral lines;* a "line" in a spectrum, no matter what technique is used to record the spectrum, is taken to mean a particular wavelength component.

In general, gratings may produce several images of spectral lines, corresponding to $m = \pm 1, \pm 2, \ldots,$ in

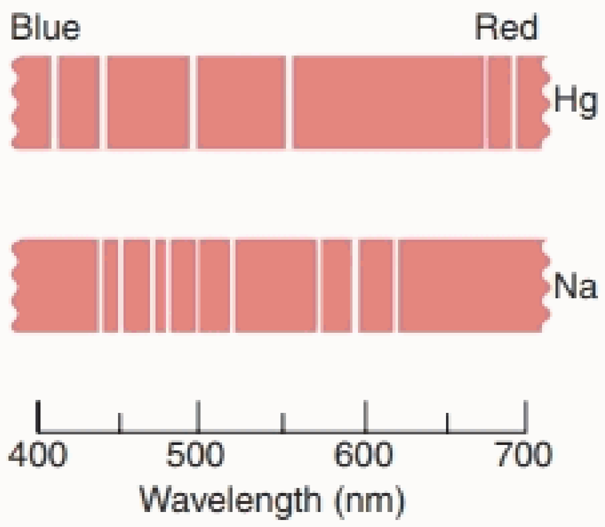


FIGURE 43-9. Examples of spectra of visible light emitted by gases of sodium (Na) and mercury (Hg).

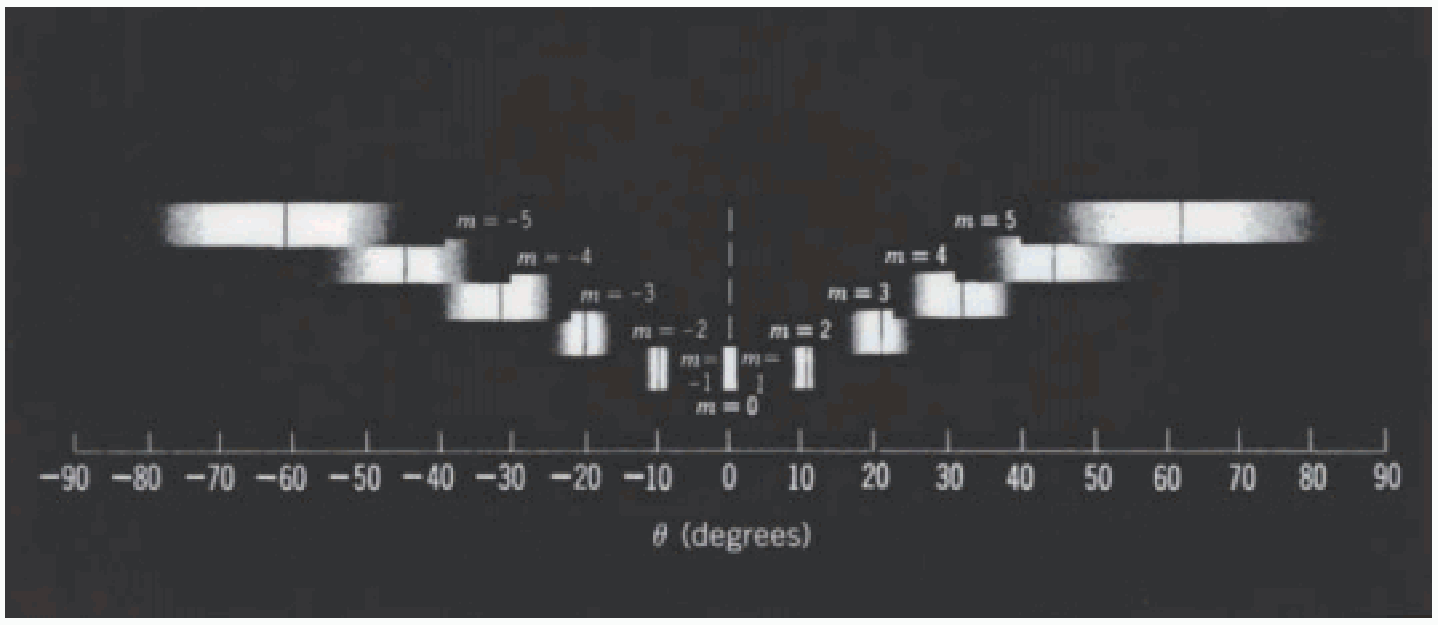


FIGURE 43-10. The spectrum of white light as viewed in a grating spectroscope such as that of Fig. 43-8. The different orders, identified by the index *m*, are shown separated vertically for clarity. As actually viewed, they would not be so displaced. The central line in each order corresponds to a wavelength of 550 nm. Diffraction gratings in common use today are designed to concentrate the intensity of the light in a particular order, and they do not show the symmetrical patterns illustrated here.

Eq. 43-1, and they can also separate wavelengths that are distributed continuously (as in Fig. 43-10) rather than as sharp spectral lines. The light from a hot, glowing object such as a lamp filament or the Sun gives a continuous spectrum. The Sun's spectrum also contains sharp spectral lines, which appear as dark lines superimposed on the continuous spectrum. These lines are caused by absorption of light by atoms of elements in the atmosphere surrounding the Sun. The element helium (from the Greek word *helios,* meaning the Sun) was discovered from an analysis of these lines.

Light can also be analyzed into its component wavelengths if the grating in Fig. 43-8 is replaced by a prism. In a *prism spectrograph,* each wavelength in the incident beam is deflected through a definite angle θ, determined by the index of refraction of the prism material for that wavelength. Curves such as Fig. 39-11, which gives the index of refraction of fused quartz as a function of wavelength, show that the shorter the wavelength, the larger is the angle of deflection θ. Such curves vary from substance to substance and must be found by measurement. Prism instruments are not adequate for accurate *absolute* measurements of wavelength because the index of refraction of the prism material at the wavelength in question is usually not known precisely enough. Both prism and grating instruments make accurate *comparisons* of wavelength, using a suitable comparison spectrum such as that shown in Fig. 43-9, in which careful absolute determinations have been made of the wavelengths of the spectral lines.

SAMPLE PROBLEM 43-2. A diffraction grating has 10^4 rulings uniformly spaced over 25.0 mm. It is illuminated at normal incidence by yellow light from a sodium vapor lamp. This light contains two closely spaced lines (the well-known *sodium doublet*) of wavelengths 589.00 and 589.59 nm. (*a*) At what angle will the first-order maximum occur for the first of these wavelengths? (*b*) What is the angular separation between the first-order maxima for these lines?

Solution (*a*) The grating spacing d is 2500 nm. The first-order maximum corresponds to $m = 1$ in Eq. 43-1. We thus have

$$\theta = \sin^{-1}\left(\frac{m\lambda}{d}\right) = \sin^{-1}\left(\frac{(1)(589\text{ nm})}{2500\text{ nm}}\right) = 13.6^\circ.$$

(*b*) The straightforward way to find the angular separation is to repeat the calculation of part (*a*) for $\lambda = 589.59$ nm and to subtract the two angles. A difficulty, which can best be appreciated by doing the calculation, is that we must carry a large number of significant figures to obtain a meaningful value for the difference between the angles. To calculate the difference in angular positions *directly,* let us solve Eq. 43-1 for $\sin\theta$ and differentiate the result, treating θ and λ as variables:

$$\sin\theta = \frac{m\lambda}{d}$$

$$\cos\theta\, d\theta = \frac{m}{d}\, d\lambda.$$

If the wavelengths are close enough together, as in this case, $d\lambda$ can be replaced by $\Delta\lambda$, the actual wavelength difference; $d\theta$ then becomes $\Delta\theta$, the quantity we seek. This gives

$$\Delta\theta = \frac{m\,\Delta\lambda}{d\cos\theta} = \frac{(1)(0.59\text{ nm})}{(2500\text{ nm})(\cos 13.6^\circ)}$$

$$= 2.4 \times 10^{-4}\text{ rad} = 0.014^\circ.$$

Note that although the wavelengths involve five significant figures, our calculation, done this way, involves only two or three, with consequent reduction in numerical manipulation.

43-3 DISPERSION AND RESOLVING POWER

The ability of a grating to produce spectra that permit precise measurement of wavelengths is determined by two intrinsic properties of the grating: (1) the separation $\Delta\theta$ between spectral lines that differ in wavelength by a small amount $\Delta\lambda$ and (2) the width or sharpness of the lines.

In Sample Problem 43-2, we calculated the angular separation between the closely spaced lines of the yellow sodium doublet, for which $\Delta\lambda = 0.59$ nm. We found in that case a separation of $\Delta\theta = 0.014^\circ$ between the first-order principal maxima of these lines. The angular separation $\Delta\theta$ per unit wavelength interval $\Delta\lambda$ is called the *dispersion D* of the grating, or

$$D = \frac{\Delta\theta}{\Delta\lambda}. \tag{43-7}$$

For lines of nearly equal wavelengths to appear as widely separated as possible, we would like our grating to have the largest possible dispersion.

To see what physical property of the grating determines its dispersion, we differentiate Eq. 43-1 ($d \sin \theta = m\lambda$), treating θ and λ as variables, which gives

$$d \cos \theta \, d\theta = m \, d\lambda,$$

or, in terms of small differences instead of differentials,

$$d \cos \theta \, \Delta\theta = m \, \Delta\lambda. \tag{43-8}$$

The dispersion D is given by $\Delta\theta/\Delta\lambda$, or

$$D = \frac{m}{d \cos \theta}. \tag{43-9}$$

The dispersion increases as the spacing between the slits decreases. We can also increase the dispersion by working at higher order (large m), as Fig. 43-10 illustrates. Note that the dispersion does not depend on the number of rulings.

Resolving Power

If a grating produces lines of large width, then the maxima of spectral lines of closely spaced wavelengths may overlap, making it difficult to determine whether such lines have one or more components and to measure the wavelengths of the lines to high precision. We therefore want to select a grating that produces the narrowest possible lines.

We obtain a reasonable measure of the ability to resolve nearby lines of different wavelengths by applying Rayleigh's criterion (see Section 42-4): if the maximum of one line falls on the first minimum of its neighbor, we should be able to resolve the lines. In Section 43-1, we defined the width of a spectra line in just that way, as the angular interval $\delta\theta$ from the maximum to the first minimum. The limit of resolution of the grating occurs when two lines in the spectrum are separated by a wavelength interval $\Delta\lambda$ such that the difference $\delta\theta$ between their angular positions is given by Eq. 43-6. We define the *resolving power R* of the grating as

$$R = \frac{\lambda}{\Delta\lambda}. \tag{43-10}$$

If the lines are to be narrow ($\delta\theta$ is small), then the corresponding wavelength interval $\Delta\lambda$ must be small, and the resolving power must be large. We should therefore choose a grating with the largest R.

To find the physical property of the grating that determines R, let us solve Eq. 43-8 for the spacing $\Delta\theta$ between nearby lines and (using Rayleigh's criterion) set this result equal to the width $\delta\theta$ of the line, given by Eq. 43-6 as the spacing between the maximum and first minimum. This gives

$$\frac{m \, \Delta\lambda}{d \cos \theta} = \frac{\lambda}{Nd \cos \theta},$$

and solving for R ($=\lambda/\Delta\lambda$) gives

$$R = Nm. \tag{43-11}$$

TABLE 43-1 **Properties of Three Gratings**[a]

Grating	N	d (nm)	θ	R	D (10^{-4} rad/nm)
A	5,000	10,000	2.9°	5,000	1.0
B	5,000	5,000	5.7°	5,000	2.0
C	10,000	10,000	2.9°	10,000	1.0

[a] For $\lambda = 500$ nm and $m = 1$.

The resolving power, like the dispersion, increases with the order number. Unlike the dispersion, R depends on the number of lines N but is independent of their separation d. To maximize the resolving power, we choose a grating with the largest number of lines. For a given slit spacing d, the grating with the greatest total width has the greatest resolving power (that is, it produces the sharpest spectral lines).

Dispersion and resolving power measure different aspects of a diffraction grating's ability to produce cleanly separated lines. Consider, for example, three gratings A, B, and C whose properties are listed in Table 43-1. Suppose that the gratings are illuminated with light consisting of a doublet of lines at 500 nm separated by an interval $\Delta\lambda = 0.10$ nm. We have chosen the properties of grating A such that the two lines of the doublet in the first-order maximum are just at the limit of resolution; that is, the maximum of one line falls on the minimum of the other, as shown in Fig. 43-11*a*. Grating B has twice the dispersion of A but the same resolving power, and it produces the spectrum shown in Fig. 43-11*b*. In effect, all angular intervals are scaled by

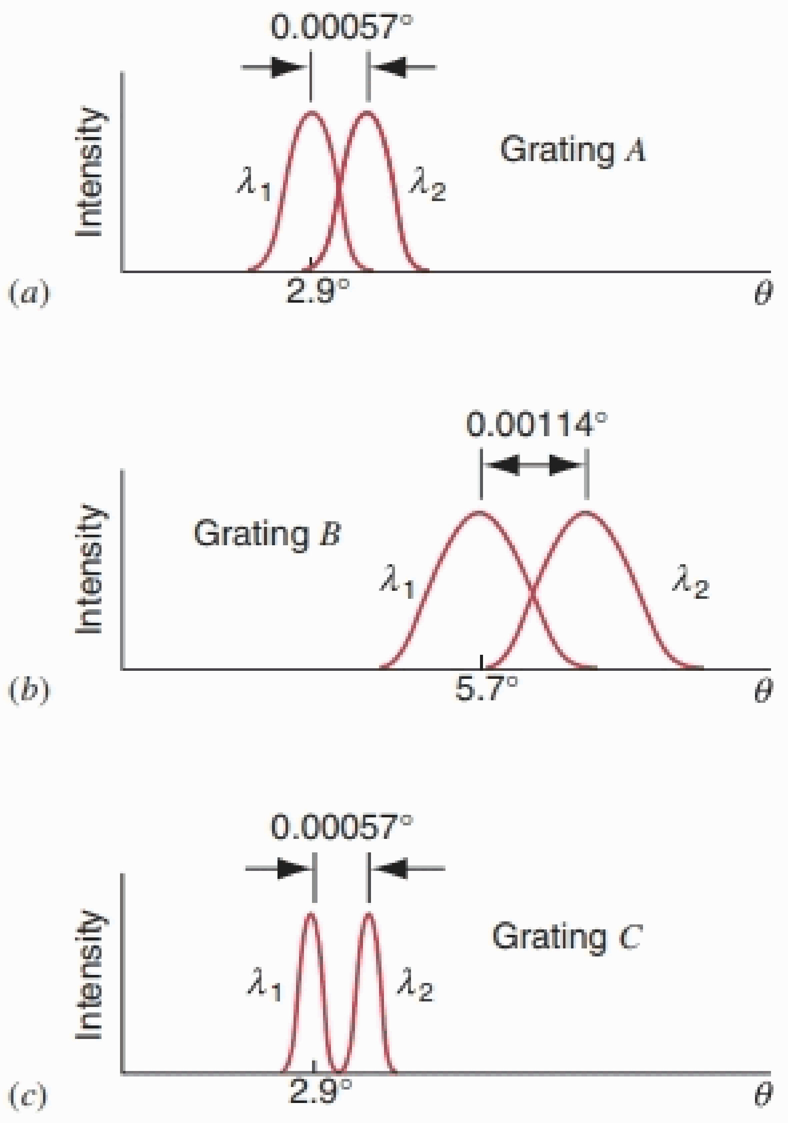


FIGURE 43-11. The intensity pattern of two lines at $\lambda = 500$ nm separated by $\Delta\lambda = 0.10$ nm, produced by the three gratings of Table 43-1. Grating *B* has the largest dispersion and grating *C* the largest resolving power.

a factor of 2, including the angular width and angular separation of the peaks. If our measurement with grating A had been limited by our ability to determine small angular intervals, changing to grating B would improve the measurement.

Grating C has twice the resolving power of A but the same dispersion. The peaks in Fig. 43-11c appear with the same angular separation as those in Fig. 43-11a, but with smaller widths. The maximum of one peak now clearly falls outside the first minimum of the other, and the two lines are more clearly distinguished from one another using grating C.

The total widths of the three gratings, equal to the product Nd, are 50 mm for grating A, 25 mm for grating B, and 100 mm for grating C. Note from Fig. 43-11 that the peak widths depend inversely on the grating width, as suggested by Eq. 43-6.

SAMPLE PROBLEM 43-3. A grating has 9600 lines uniformly spaced over a width $W = 3.00$ cm and is illuminated by light from a mercury vapor discharge. (*a*) What is the expected dispersion, in the third order, in the vicinity of the intense green line ($\lambda = 546$ nm)? (*b*) What is the resolving power of this grating in the fifth order?

Solution (*a*) The grating spacing d is given by

$$d = \frac{W}{N} = \frac{3.00 \times 10^{-2}\text{ m}}{9600} = 3125\text{ nm}.$$

We must find the angle θ at which the line in question occurs. From Eq. 43-1, we have

$$\theta = \sin^{-1}\left(\frac{m\lambda}{d}\right) = \sin^{-1}\left(\frac{(3)(546\text{ nm})}{3125\text{ nm}}\right) = 31.6°.$$

We can now calculate the dispersion. From Eq. 43-9,

$$\begin{aligned} D &= \frac{m}{d\cos\theta} = \frac{3}{(3125\text{ nm})(\cos 31.6°)} \\ &= 1.13 \times 10^{-3}\text{ rad/nm} \\ &= 0.0646°/\text{nm} = 3.87\text{ arc min/nm}. \end{aligned}$$

(*b*) From Eq. 43-11,

$$R = Nm = (9600)(5) = 4.80 \times 10^4.$$

Thus, near $\lambda = 546$ nm and in fifth order, a wavelength difference given by (see Eq. 43-10)

$$\Delta\lambda = \frac{\lambda}{R} = \frac{546\text{ nm}}{4.80 \times 10^4} = 0.011\text{ nm}$$

can be resolved.

SAMPLE PROBLEM 43-4. A diffraction grating has 1.20×10^4 rulings uniformly spaced over a width $W = 2.50$ cm. It is illuminated at normal incidence by yellow light from a sodium vapor lamp. This light contains two closely spaced lines of wavelengths 589.00 and 589.59 nm. (*a*) At what angle does the first-order maximum occur for the first of these wavelengths? (*b*) What is the angular separation between these two lines (in first order)? (*c*) How close in wavelength can two lines be (in first order) and still be resolved by this grating? (*d*) How many rulings can a grating have and just resolve the sodium doublet lines?

Solution (*a*) The grating spacing d is given by

$$d = \frac{W}{N} = \frac{2.50 \times 10^{-2}\text{ m}}{1.20 \times 10^4} = 2083\text{ nm}.$$

The first-order maximum corresponds to $m = 1$ in Eq. 43-1. We thus have

$$\theta = \sin^{-1}\left(\frac{m\lambda}{d}\right) = \sin^{-1}\left(\frac{(1)(589.00\text{ nm})}{2083\text{ nm}}\right) = 16.4°.$$

(*b*) Here the *dispersion* of the grating comes into play. From Eq. 43-9, the dispersion is

$$\begin{aligned} D &= \frac{m}{d\cos\theta} = \frac{1}{(2083\text{ nm})(\cos 16.4°)} \\ &= 5.00 \times 10^{-4}\text{ rad/nm}. \end{aligned}$$

From Eq. 43-7, the defining equation for dispersion, we have

$$\begin{aligned} \Delta\theta &= D\,\Delta\lambda \\ &= (5.00 \times 10^{-4}\text{ rad/nm})(589.59\text{ nm} - 589.00\text{ nm}) \\ &= 2.95 \times 10^{-4}\text{ rad} = 0.0169° = 1.01\text{ arc min}. \end{aligned}$$

As long as the grating spacing d remains fixed, this result holds no matter how many lines there are in the grating.

(*c*) Here the *resolving power* of the grating comes into play. From Eq. 43-11, the resolving power is

$$R = Nm = (1.20 \times 10^4)(1) = 1.20 \times 10^4.$$

From Eq. 43-10, the defining equation for resolving power, we have

$$\Delta\lambda = \frac{\lambda}{R} = \frac{589\text{ nm}}{1.20 \times 10^4} = 0.049\text{ nm}.$$

This grating can easily resolve the two sodium lines, which have a wavelength separation of 0.59 nm. Note that this result depends only on the number of grating rulings and is independent of d, the spacing between adjacent rulings.

(*d*) From Eq. 43-10, the defining equation for R, the grating must have a resolving power of

$$R = \frac{\lambda}{\Delta\lambda} = \frac{589\text{ nm}}{0.59\text{ nm}} = 998.$$

From Eq. 43-11, the number of rulings needed to achieve this resolving power (in first order) is

$$N = \frac{R}{m} = \frac{998}{1} = 998\text{ rulings}.$$

Since the grating has about 12 times as many rulings as this, it can easily resolve the sodium doublet lines, as we have already shown in part (*c*).

43-4 X-RAY DIFFRACTION

X rays are electromagnetic radiation with wavelengths of the order of 0.1 nm (compared with 500 nm for a typical wavelength of visible light). Figure 43-12 shows how x rays are produced when electrons from a heated filament

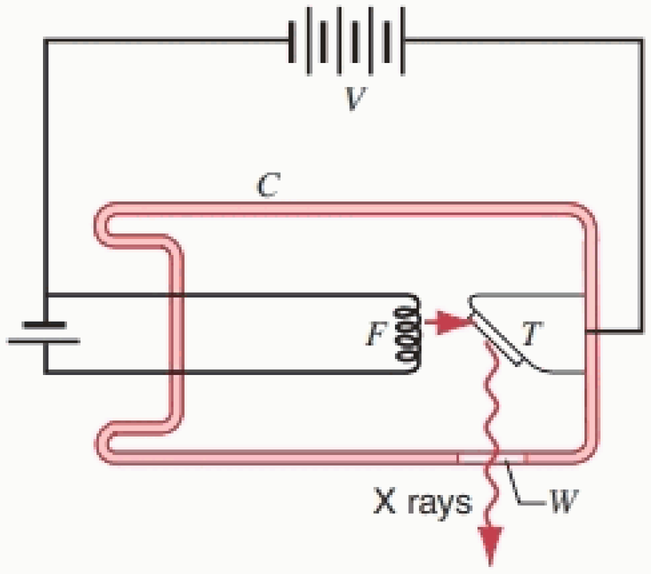


FIGURE 43-12. X rays are generated when electrons from heated filament F, accelerated through a potential difference V, strike a metal target T in the evacuated chamber C. Window W is transparent to x rays.

F are accelerated by a potential difference V and strike a metal target.

For such small wavelengths, a standard optical diffraction grating, as normally employed, cannot be used. For $\lambda = 0.10$ nm and $d = 3000$ nm, for example, Eq. 43-1 shows that the first-order maximum occurs at

$$\theta = \sin^{-1}\left(\frac{m\lambda}{d}\right) = \sin^{-1}\left(\frac{(1)(0.10\text{ nm})}{3 \times 10^3\text{ nm}}\right)$$
$$= 0.0019°.$$

This is too close to the central maximum to be practical. A grating with $d \approx \lambda$ is desirable, but, because x-ray wavelengths are about equal to atomic diameters, such gratings cannot be constructed mechanically.

In 1912, it occurred to physicist Max von Laue that a crystalline solid, consisting as it does of a regular array of atoms, might form a natural three-dimensional "diffraction grating" for x rays. Figure 43-13 shows that if a collimated beam of x rays, continuously distributed in wavelength, is allowed to fall on a crystal, such as sodium chloride, intense beams (corresponding to constructive interference from the many diffracting centers of which the crystal is made up) appear in certain sharply defined directions. If these beams fall on a photographic film, they form an assembly of "Laue spots." Figure 43-14, which shows an actual example of these spots, demonstrates that the hypothesis of Laue is indeed correct. The atomic arrangements in the crystal can be deduced from a careful study of the positions and intensities of the Laue spots in much the same way that we might deduce the structure of an optical grating (that is, the detailed profile of its slits) by a study of the positions and intensities of the lines in the interference pattern. Other experimental arrangements have supplanted the Laue technique to a considerable extent today, but the principle remains unchanged (see Question 23).

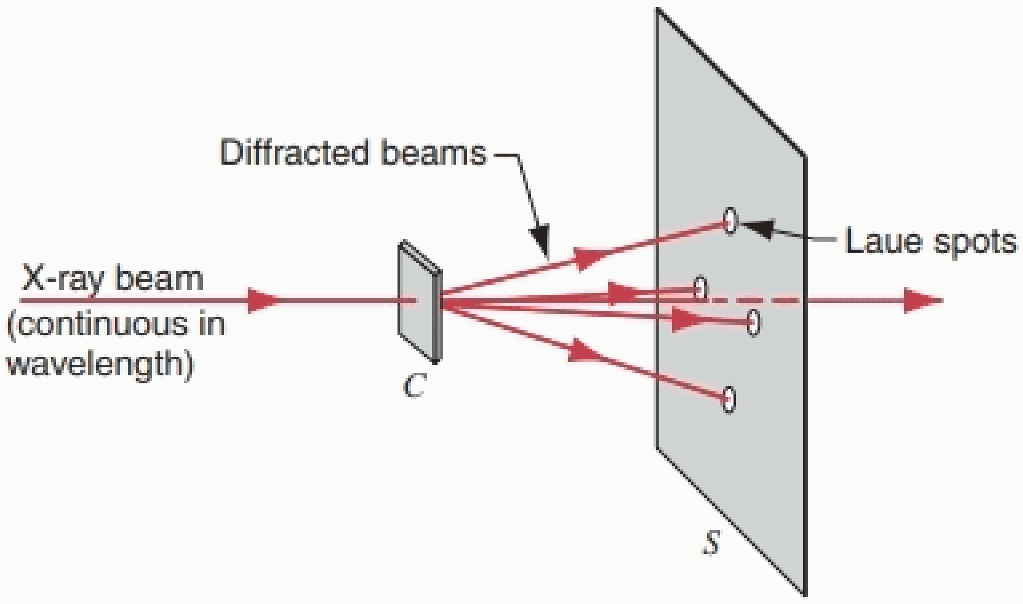


FIGURE 43-13. A beam of x rays strikes a crystal C. Strong diffracted beams appear in certain directions, forming a Laue pattern on the photographic film S.

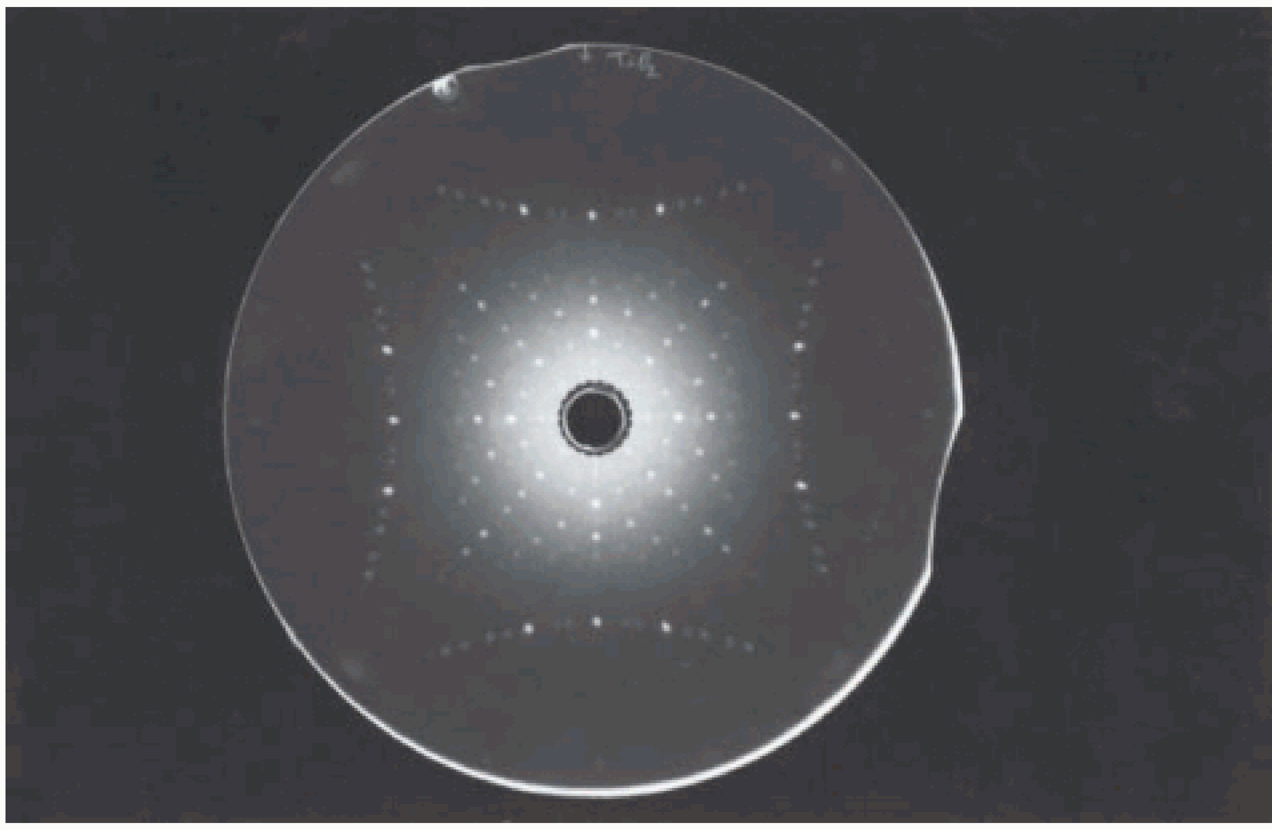


FIGURE 43-14. A Laue x-ray diffraction pattern from a crystal of sodium chloride.

Figure 43-15 shows how sodium and chlorine atoms (strictly, Na^+ and Cl^- ions) are stacked to form a crystal of sodium chloride. This pattern, which has *cubic* symmetry,

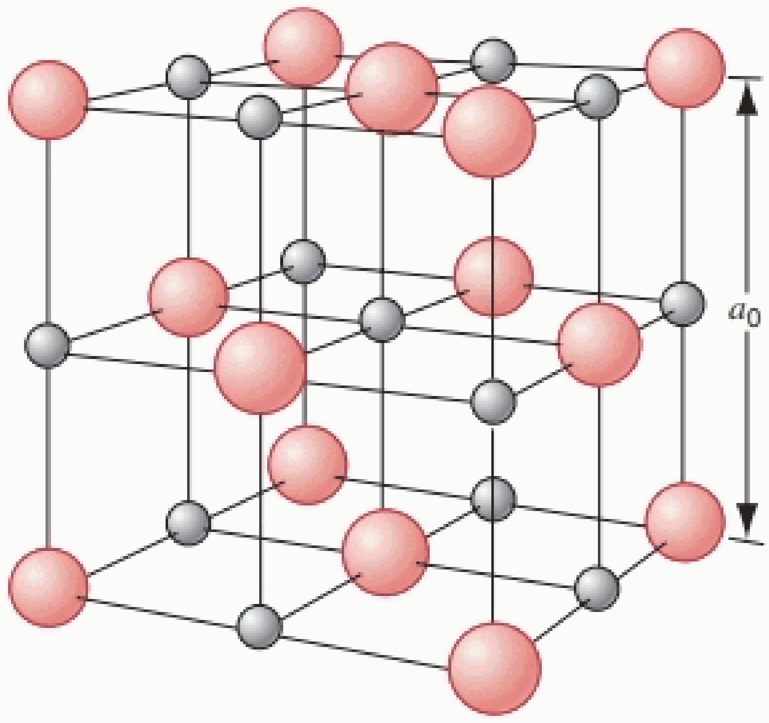


FIGURE 43-15. A model of a sodium chloride crystal, showing how the sodium ions Na^+ (small spheres) and chloride ions Cl^- (large spheres) are stacked in the unit cell, whose edge a_0 has the length 0.563 nm.

is one of the many possible atomic arrangements exhibited by solids. The model represents the *unit cell* for sodium chloride. This is the smallest unit from which the crystal may be built up by repetition in three dimensions. You should verify that no smaller assembly of atoms possesses this property. For sodium chloride, the length a_0 of the cube edge of the unit cell is 0.563 nm.

Each unit cell in sodium chloride has four sodium ions and four chlorine ions associated with it. In Fig. 43-15, the sodium ion in the center belongs entirely to the cell shown. Each of the other twelve sodium ions shown is shared with three adjacent unit cells so that each contributes one-fourth of an ion to the cell under consideration. The total number of sodium ions is then $1 + \frac{1}{4}(12) = 4$. By similar reasoning you can show that although there are fourteen chlorine ions in Fig. 43-15, only four are associated with the unit cell shown.

The unit cell is the fundamental repetitive diffracting unit in the crystal, corresponding to the slit (and its adjacent opaque strip) in the optical diffraction grating of Fig. 43-1. Figure 43-16*a* shows a particular plane in a sodium chloride crystal. If each unit cell intersected by this plane is represented by a small cube, Fig. 43-16*b* results. You may imagine each of these figures extended indefinitely in three dimensions.

Let us treat each small cube in Fig. 43-16*b* as an elementary diffracting center, corresponding to a slit in an optical grating. The *directions* (but not the intensities) of all the diffracted x-ray beams that can emerge from a sodium chloride crystal (for a given x-ray wavelength and a given orientation of the incident beam) are determined by the geometry of this three-dimensional lattice of diffracting centers. In exactly the same way, the *directions* (but not the intensities) of all the diffracted beams that can emerge from a particular optical grating (for a given wavelength and orientation of the incident beam) are determined only by the geometry of the grating—that is, by the grating spacing *d*. Representing the unit cell by what is essentially a point, as in Fig. 43-16*b*, corresponds to representing the slits in a diffraction grating by lines, as we did in discussing the double-slit experiment in Section 41-2.

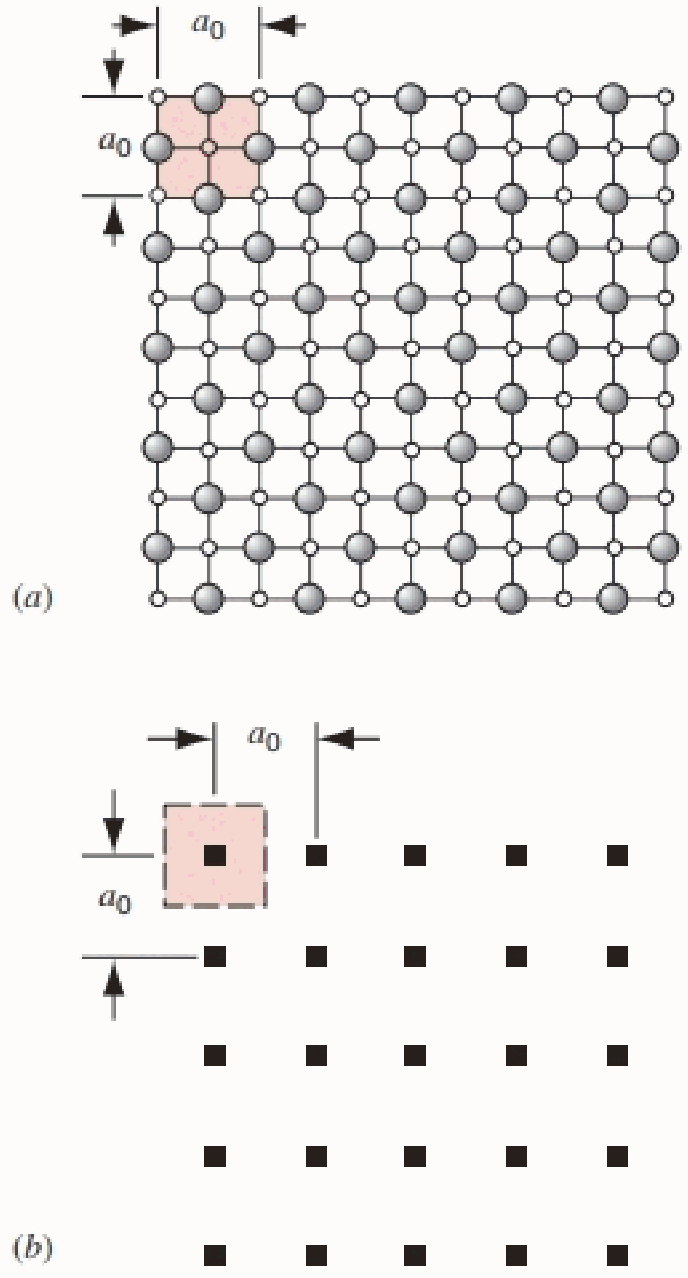


FIGURE 43-16. (*a*) A plane through a crystal of NaCl, showing the Na and Cl ions. (*b*) The corresponding unit cells in this section. Each cell is represented by a small black square.

The *intensities* of the lines from an optical diffraction grating depend on the diffracting characteristics of a single slit, determined in particular by the slit width *a*; see, for example, Fig. 43-1 for a set of slits. The characteristics of actual optical gratings are determined by the profile of the grating rulings.

In exactly the same way, the *intensities* of the diffracted beams emerging from a crystal depend on the diffracting characteristics of the unit cell. Fundamentally, the x rays are diffracted by electrons, diffraction by nuclei being negligible in most cases. Thus the diffracting characteristics of a unit cell depend on how the electrons are distributed throughout the volume of the cell. By studying the *directions* of diffracted x-ray beams, we can learn the basic symmetry of the crystal. By studying the *intensities* we can learn how the electrons are distributed in a unit cell. Figure 43-17 shows an example of this technique.

Bragg's Law

Bragg's law predicts the conditions under which diffracted x-ray beams from a crystal are possible. In deriving the law, we ignore the structure of the unit cell, which is related only to the intensities of these beams. The dashed sloping lines in Fig. 43-18*a* represent the intersection with the plane of the figure of an arbitrary set of planes passing through the elementary diffracting centers. The perpendicular distance between adjacent planes is *d*. Many other such families of planes, with different *interplanar spacings,* can be defined.

Figure 43-18*b* shows an incident wave striking the *family* of planes, with the incident rays making an angle θ with the plane.* For a single plane, mirror-like "reflection" occurs for *any* value of θ. To have a constructive interference in the beam diffracted from the entire family of planes in the direction θ, the rays from the separate planes must reinforce each other. This means that the path difference for rays from adjacent planes (*abc* in Fig. 43-18*b*) must be an integral number of wavelengths or

$$2d \sin \theta = m\lambda \qquad m = 1, 2, 3, \ldots . \qquad (43\text{-}12)$$

---

* In x-ray diffraction it is customary to specify the direction of a wave by giving the angle between the ray and the plane (the *glancing angle*) rather than the angle between the ray and the normal.

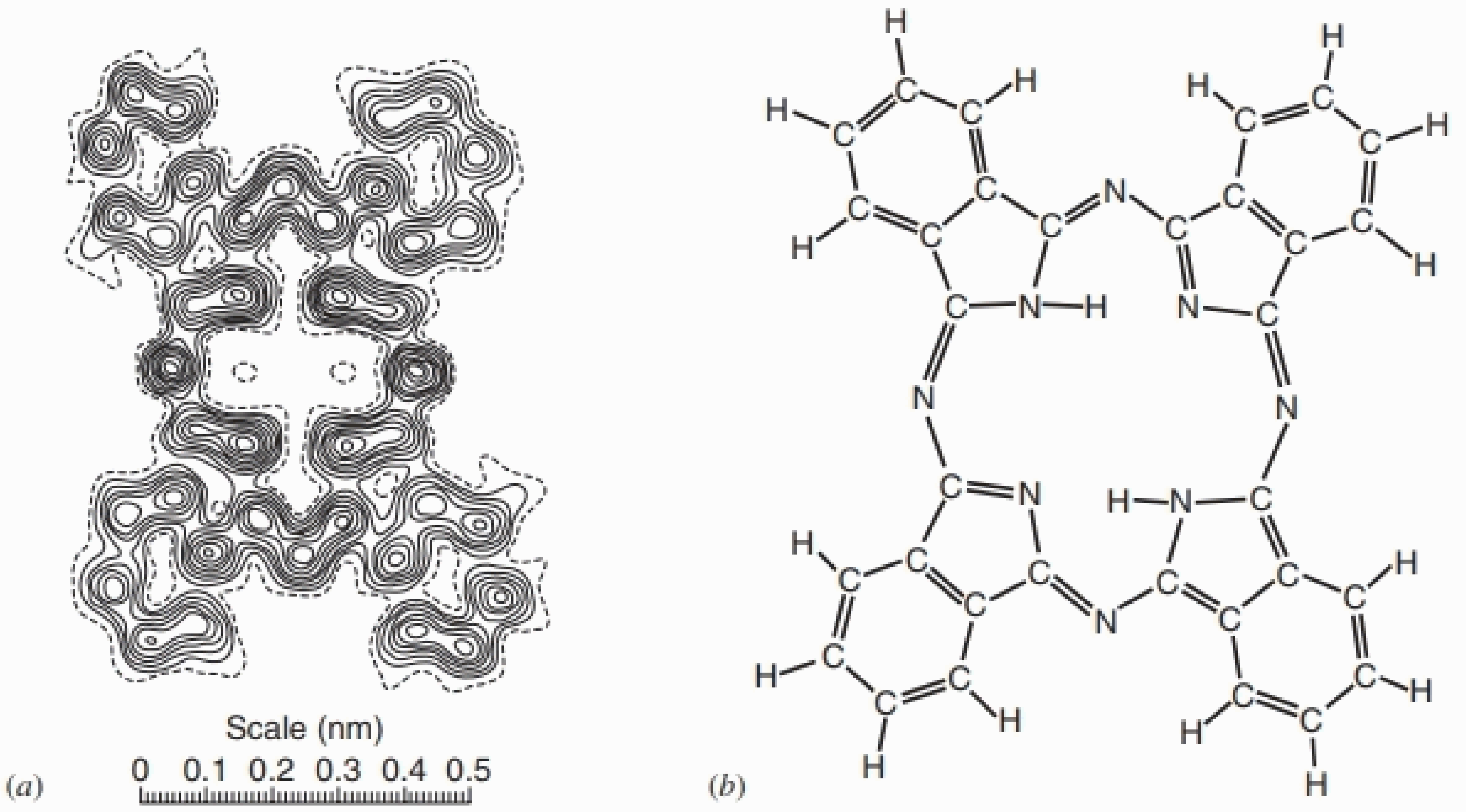


FIGURE 43-17. (*a*) Electron density contours for phthalocyanine ($C_{32}H_{18}N_8$) determined from the intensity distribution of scattered x rays. The dashed curves represent a density of one electron per 0.01 nm^2, and each adjacent curve represents an increase of one electron per 0.01 nm^2. (*b*) A structural representation of the molecule. Note that the greatest electron density occurs in (*a*) near the N atoms, which have the largest number of electrons (7). Note also that the H atoms, which contain only a single electron, are not prominent in (*a*).

This relation is called *Bragg's law* after W. L. Bragg, who first derived it. The quantity d in this equation (the interplanar spacing) is the perpendicular distance between the planes. For the planes of Fig. 43-18*a* we see that d is related to the unit cell dimension a_0 by

$$d = \frac{a_0}{\sqrt{5}}. \tag{43-13}$$

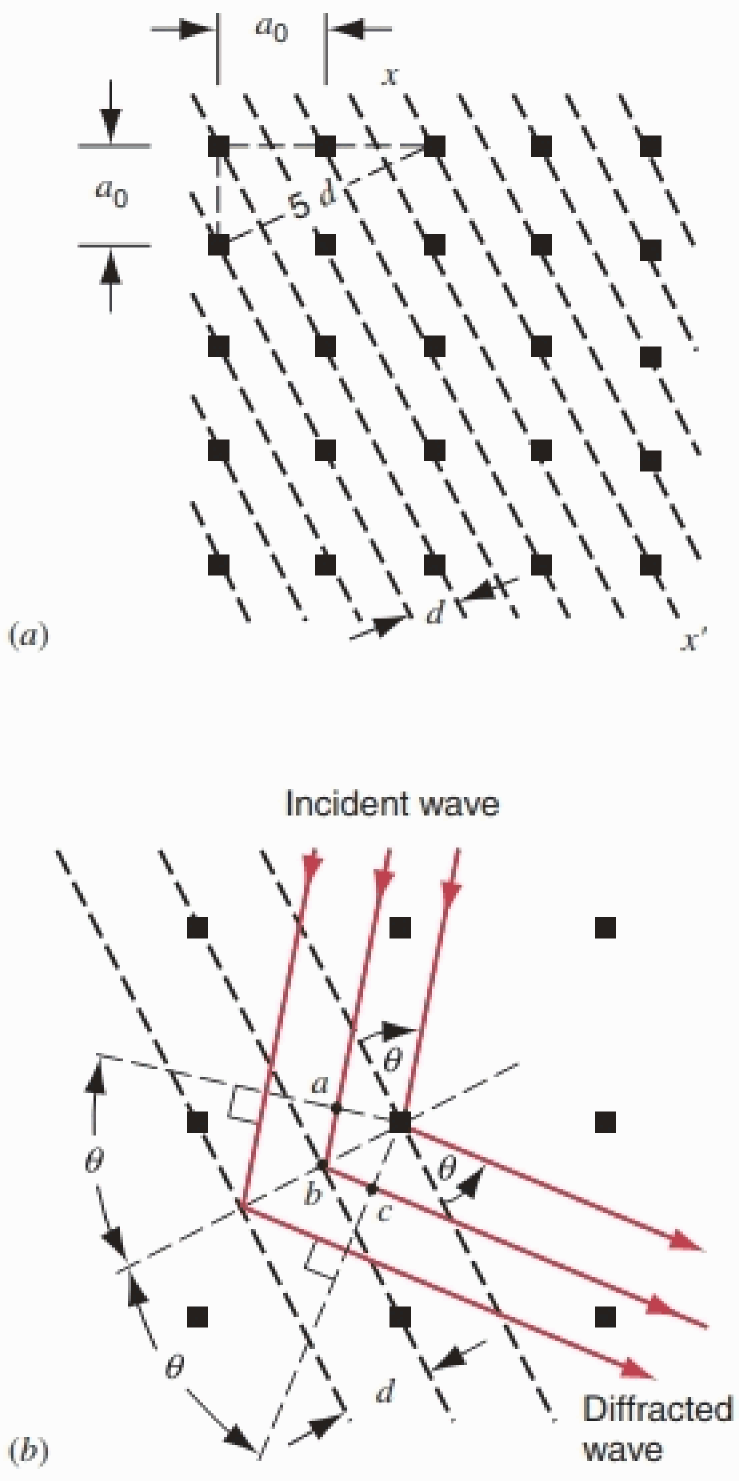


FIGURE 43-18. (*a*) A section through the NaCl lattice of Fig. 43-16. The dashed lines represent an arbitrary set of parallel planes connecting unit cells. The interplanar spacing is *d*. (*b*) An incident beam falls on a set of planes. A strong diffracted beam will be observed if Bragg's law is satisfied.

If an incident monochromatic x-ray beam falls at an arbitrary angle θ on a particular set of atomic planes, a diffracted beam will *not* result because Eq. 43-12 will not, in general, be satisfied. If the incident x rays are *continuous* in wavelength, diffracted beams will result when wavelengths given by

$$\lambda = \frac{2d \sin\theta}{m} \qquad m = 1, 2, 3, \ldots$$

are present in the incident beam (see Eq. 43-12).

X-ray diffraction is a powerful tool for studying both x-ray spectra and the arrangement of atoms in crystals. To study the spectrum of an x-ray source, a particular set of crystal planes, having a known spacing d, is chosen. Diffraction from these planes locates different wavelengths at different angles. A detector that can discriminate one angle from another can be used to determine the wavelength of radiation reaching it. On the other hand, we can study the crystal itself, using a monochromatic x-ray beam to determine not only the spacings of various crystal planes but also the structure of the unit cell. The DNA molecule and many other equally complex structures have been mapped by x-ray diffraction methods.

SAMPLE PROBLEM 43-5. At what angles must an x-ray beam with $\lambda = 0.110$ nm fall on the family of planes represented in Fig. 43-18*b* if a diffracted beam is to exist? Assume the material to be sodium chloride ($a_0 = 0.563$ nm).

Solution The interplanar spacing d for these planes is given by Eq. 43-13, or

$$d = \frac{a_0}{\sqrt{5}} = \frac{0.563\text{ nm}}{\sqrt{5}} = 0.252\text{ nm}.$$

Equation 43-12 gives

$$\theta = \sin^{-1}\left(\frac{m\lambda}{2d}\right) = \sin^{-1}\left(\frac{(m)(0.110\text{ nm})}{(2)(0.252\text{ nm})}\right).$$

Diffracted beams are possible for $\theta = 12.6°$ ($m = 1$), $\theta = 25.9°$ ($m = 2$), $\theta = 40.9°$ ($m = 3$), and $\theta = 60.8°$ ($m = 4$). Higher-order beams cannot exist because they require that $\sin\theta > 1$.

Actually, the unit cell in cubic crystals such as NaCl has symmetry properties that require the intensity of diffracted x-ray beams corresponding to odd values of m to be zero. (See Problem 35.) Thus the only beams that are expected are $\theta = 25.9°$ ($m = 2$) and $\theta = 60.9°$ ($m = 4$).

43-5 HOLOGRAPHY (Optional)

The light emitted by an object contains the complete information on the size and shape of the object. We can consider that information to be stored in the wavefronts of the light from the object, specifically in the variation of intensity and phase of the electromagnetic fields. If we could record this information, we could reproduce a complete three-dimensional image of the object. However, photographic films record only the intensity variations; the films are not sensitive to phase variations. It is therefore not possible to use a photographic negative to reconstruct a three-dimensional image.

One exception to this restriction occurs in the case of x-ray diffraction from a crystal. Because of the regular spacing of the atoms of a crystal, we can easily deduce the relative phases of the diffracted waves reaching the film from different atoms. This possibility was realized by W. L. Bragg, who illuminated a photographic negative of an x-ray diffraction pattern and so reconstructed the image of a crystal. In this "double diffraction" method, diffraction of radiation from a diffraction pattern gives an image of the original object. For objects whose atoms are not arranged in such a periodic array, this simple method of image reconstruction does not work.

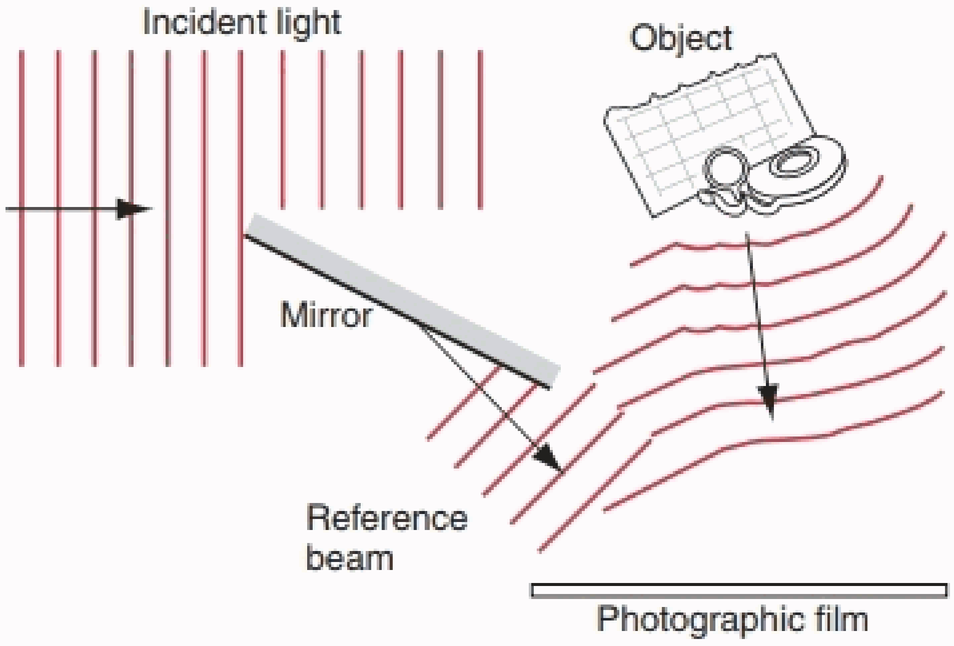


FIGURE 43-19. Apparatus for producing holograms. A portion of the beam from a source of coherent light (a laser, for instance) illuminates an object. The light diffracted by the object interferes on the film with a portion of the original beam, which serves as the reference.

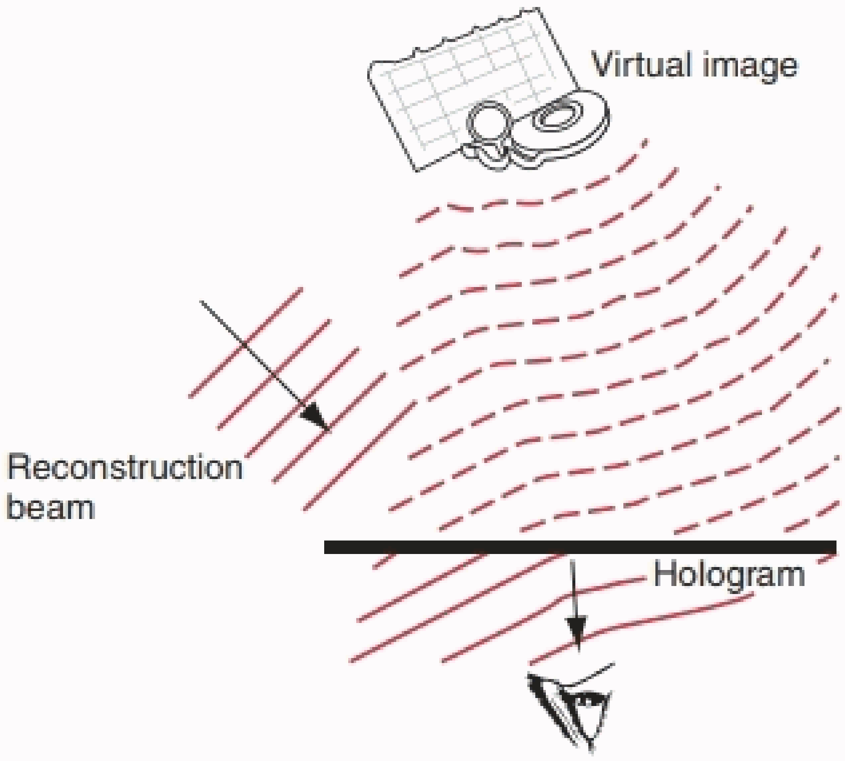


FIGURE 43-20. To view a hologram, it is illuminated with light identical to the reference beam. A three-dimensional virtual image can be seen, at the location of the original object.

A scheme for recording the intensity *and* phase of the waves from objects was developed in 1948 by Dennis Gabor, who was awarded the 1971 Nobel Prize in physics for this discovery. This type of image formation is called *holography,* from the Greek words meaning "entire picture," and the image is called a *hologram.* The procedure is illustrated in Fig. 43-19. A wave diffracted from an object interferes on the photographic film with a reference wave. The interference between the two waves serves as the means for storing on the film information on the phase of the wave from the object. When the photographic image is viewed using light identical with the reference beam, a three-dimensional virtual image of the original object is reconstructed (Fig. 43-20). A second image (a real image), not shown in Fig. 43-20, is also produced by the hologram.

Because the film is illuminated uniformly by the diffracted light from the object and the reference beam,

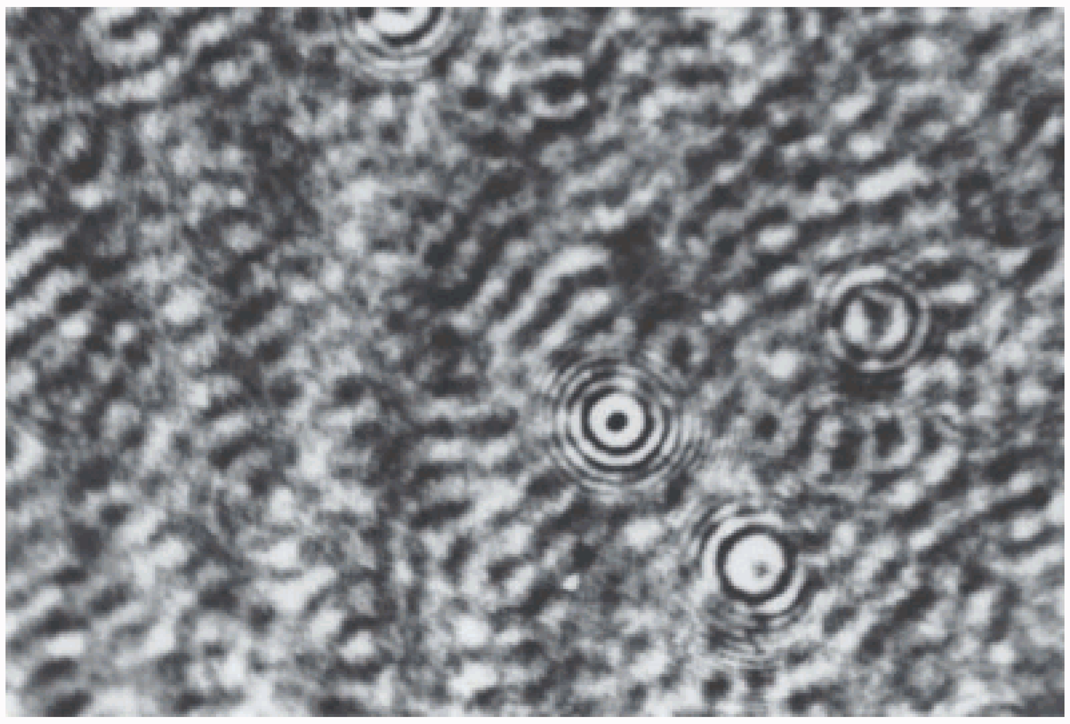


FIGURE 43-21. A close-up view of a hologram, showing the interference pattern.

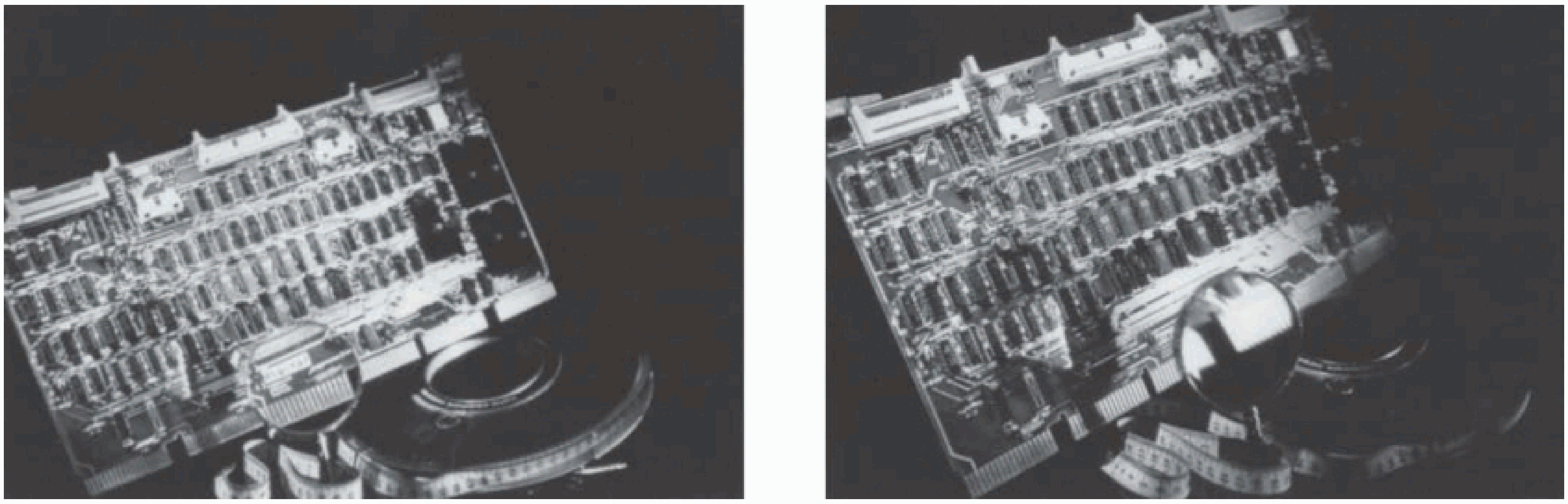


FIGURE 43-22. Two different views of the same hologram, taken from different directions. Note the relative movement of the objects in the images.

every piece of the film contains the information necessary to reproduce the three-dimensional image. The hologram itself (Fig. 43-21) shows only the interference fringes; in general, it is necessary to use a suitable monochromatic and coherent beam to reconstruct the image. For this reason, active development of holography did not occur until the early 1960s, when lasers became commonly available.

Some holograms can be viewed in white light. White-light holograms use a thick photographic emulsion, in which light is reflected by successive layers of grains in the film. Constructive interference occurs in the reflected light for the wavelength of the original reference beam, and destructive interference occurs for other wavelengths. By using reference beams of several different colors, a full-color image can be produced.

The hologram reconstructs a true three-dimensional image; for example, nearby objects appear "in front of" more distant objects, and by moving your head from side to side you can change the relative spatial orientation of the objects. Figure 43-22 shows two different views of the same hologram, illustrating the *parallax* effect of viewing the hologram from two different directions.

Holograms have a variety of applications in basic and applied science. For example, in producing holograms the object must be kept absolutely still while the film is exposed; any small movement would change the relative phase between the diffracted and reference beams and thereby change the interference pattern stored on the film. If a hologram is made by superimposing on the film two successive exposures of a vibrating object, such as the top or bottom plate of a violin, locations on the object that moved between the exposures by an integral number of wavelengths will show constructive interference, while parts of the object that moved by a half-integral number of wavelengths ($\lambda/2$, $3\lambda/2$, . . .) will show destructive interference. Figure 43-23 shows an example of the use of this technique, called *holographic interferometry.*

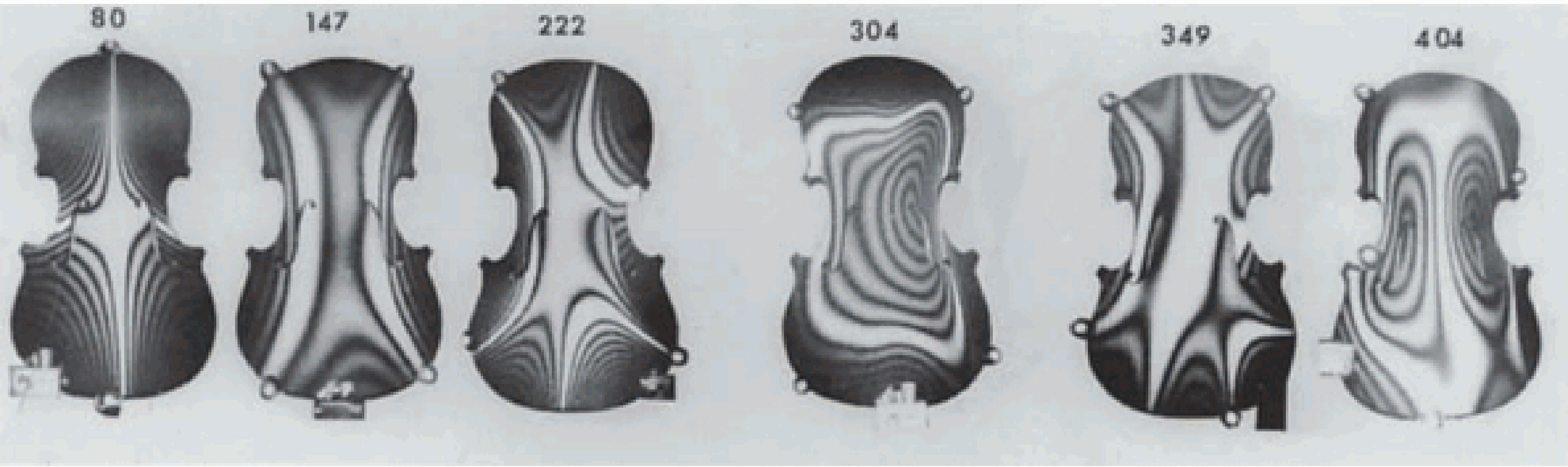


FIGURE 43-23. A holographic interference pattern of a top violin plate vibrating at different frequencies. The frequencies (in Hz) are shown above the plates.

MULTIPLE CHOICE

43-1 Multiple Slits

1. Which of the following is an effect of adding additional slits to an interference apparatus?
   (A) The bright fringes get closer together.
   (B) The bright fringes get farther apart.
   (C) The width of the bright fringes increases.
   (D) The width of the dark fringes increases.

2. Consider two gratings: one is 0.5 cm wide with 5000 lines/cm, the second is 1.0 cm wide with 2500 lines/cm. Assuming that both gratings are fully illuminated with identical light, which grating has the sharpest central maximum?
   (A) The first (B) The second
   (C) The gratings have the same sharpness.

3. A diffraction "grating" appropriate for sound waves is made by cutting slits in a large rubber sheet. The sheet is held in front of a speaker and a narrow central maximum is produced. The sheet is then stretched so that the slits are farther apart. How does the width of the central maximum change?
   (A) It gets wider. (B) It gets narrower.
   (C) It remains the same width.

43-2 Diffraction Gratings

4. Which of the following properties are true for *both* a prism and a diffraction grating?
   (A) The device divides white light into spectral lines at different angular positions.
   (B) The device generates more than one image of a spectral line.
   (C) The device can make accurate absolute measurements of wavelength.
   (D) The device deflects violet light more than red light.

5. The Sun's spectrum consists of several dark absorption lines superimposed on a continuous spectrum. Which of the following changes will help the dark bands to be more easily identified with a diffraction grating?
   (A) Increase the number of slits.
   (B) Increase the separation between slits.
   (C) Increase the width of the individual slits.
   (D) None of the above

43-3 Dispersion and Resolving Power

6. What is the difference between dispersion and sharpness?
   (A) Dispersion is a measure of the width of the maxima, whereas sharpness is how easy it is to see them.
   (B) Dispersion is a measure of the width of the maxima, whereas sharpness is a measure of the angle between two maxima.
   (C) Dispersion is a measure of the angle between two maxima, whereas sharpness is a measure of the width.
   (D) Dispersion and sharpness are effectively a measure of the same thing.

7. Consider the rubber sheet in Multiple-choice question 3.
   (*a*) What happens to the resolving power when the sheet is stretched?
   (A) It increases. (B) It decreases.
   (C) It remains the same.
   (*b*) What happens to the dispersion when the sheet is stretched?
   (A) It increases. (B) It decreases.
   (C) It remains the same.

8. Two monochromatic light beams of nearly equal wavelengths are incident on a grating of *N* rulings and are not quite resolvable. However, they become resolved if the number of rulings is increased. Formulas aside, is the explanation that
   (A) more light can get through the grating?
   (B) the principal maxima become more intense and hence resolvable?
   (C) the diffraction pattern is spread more and hence the wavelengths become resolved?
   (D) there are a large number of orders?
   (E) the principal maxima become narrower and hence resolvable?

43-4 X-ray Diffraction

9. In some respects, Bragg reflection differs from plane grating diffraction. Of the following statements, which one is true for Bragg reflection but not true for grating diffraction?
   (A) Two different wavelengths may be superposed.
   (B) Radiation of a given wavelength may be sent in more than one direction.
   (C) Long waves are deviated more than short waves.
   (D) There is only one grating spacing.
   (E) Diffraction maxima of a given wavelength occur only for particular angles of incidence.

43-5 Holography

QUESTIONS

1. Discuss this statement: "A diffraction grating can just as well be called an interference grating."

2. How would the spectrum of an enclosed source that is formed by a diffraction grating on a screen change (if at all) when the source, grating, and screen are all submerged in water?

3. (*a*) For what kind of waves could a long picket fence be considered a useful grating? (*b*) Can you make a diffraction grating out of parallel rows of fine wire strung closely together?

4. Could you construct a diffraction grating for sound? If so, what grating spacing is suitable for a wavelength of 0.5 m?

5. A crossed diffraction grating is ruled in two directions, at right angles to each other. Predict the pattern of light intensity on the screen if light is sent through such a grating. Is there any practical value to such a grating?

6. Suppose that, instead of a slit, a small, circular aperture were placed in the focal plane of the collimating lens in the tele-

scope of a spectrometer. What would be seen when the telescope is illuminated by sodium light? Whey then do we usually call spectra "line spectra"?

7. In a grating spectrograph, several lines having different wavelengths and formed in different orders might appear near a certain angle. How could you distinguish between their orders?

8. You are given a photograph of a spectrum on which the angular positions and the wavelengths of the spectrum lines are marked. (*a*) How can you tell whether the spectrum was taken with a prism or a grating instrument? (*b*) What information could you gather about either the prism or the grating from studying such a spectrum?

9. A glass prism can form a spectrum. Explain how. How many "orders" of spectra will a prism produce?

10. For the simple spectroscope of Fig. 43-8, show (*a*) that θ increases with λ for a grating and (*b*) that θ decreases with λ for a prism.

11. According to Eq. 43-6, the principal maxima become wider (that is, $\delta\theta$ increases) the higher the order m (that is, the larger θ becomes). According to Eq. 43-11, the resolving power becomes greater the higher the order m. Explain this apparent paradox.

12. Explain in your own words why increasing the number of slits N in a diffraction grating sharpens the maxima. Why does decreasing the wavelength do so? Why does increasing the grating spacing d do so?

13. How much information can you discover about the structure of a diffraction grating by analyzing the spectrum it forms of a monochromatic light source? Let $\lambda = 589$ nm, for example.

14. Assume that the limits of the visible spectrum are 430 and 680 nm. How would you design a grating, assuming that the incident light falls normally on it, such that the first-order spectrum barely overlaps the second-order spectrum?

15. (*a*) Why does a diffraction grating have closely spaced rulings? (*b*) Why does it have a large number of rulings?

16. The relation $R = Nm$ suggests that the resolving power of a given grating can be made as large as desired by choosing an arbitrarily high order of diffraction. Discuss this possibility.

17. Show that at a given wavelength and a given angle of diffraction the resolving power of a grafting depends only on its width $W\ (= Nd)$.

18. How would you experimentally measure (*a*) the dispersion D and (*b*) the resolving power R of a grating spectrograph?

19. For a given family of planes in a crystal, can the wavelength of incident x rays be (*a*) too large or (*b*) too small to form a diffracted beam?

20. If a parallel beam of x rays of wavelength λ is allowed to fall on a randomly oriented crystal of any material, generally no intense diffracted beams will occur. Such beams appear if (*a*) the x-ray beam consists of a continuous distribution of wavelengths rather than a single wavelength or (*b*) the specimen is not a single crystal but a finely divided powder. Explain each case.

21. Does an x-ray beam undergo refraction as it enters and leaves a crystal? Explain your answer.

22. Why can a simple cube of edge $a_0/2$ in Fig. 43-15 not be used as a unit cell for sodium chloride?

23. In Fig. 43-24*a* we show schematically the Debye–Scherrer experimental arrangement and in Figure 43-24*b* a corresponding x-ray diffraction pattern. (*a*) Keeping in mind that the Laue method uses a large single crystal and an x-ray beam continuously distributed in wavelength, explain the origin of the spots in Fig. 43-14. (Hint: Each spot corresponds to the direction of scattering from a family of planes.) (*b*) Keeping in mind that the Debye–Scherrer method uses a large number of small single crystals randomly oriented and a monochromatic beam of x rays, explain the origin of the rings. (Hint: Because the small crystals are randomly oriented, all possible angles of incidence are obtained.)

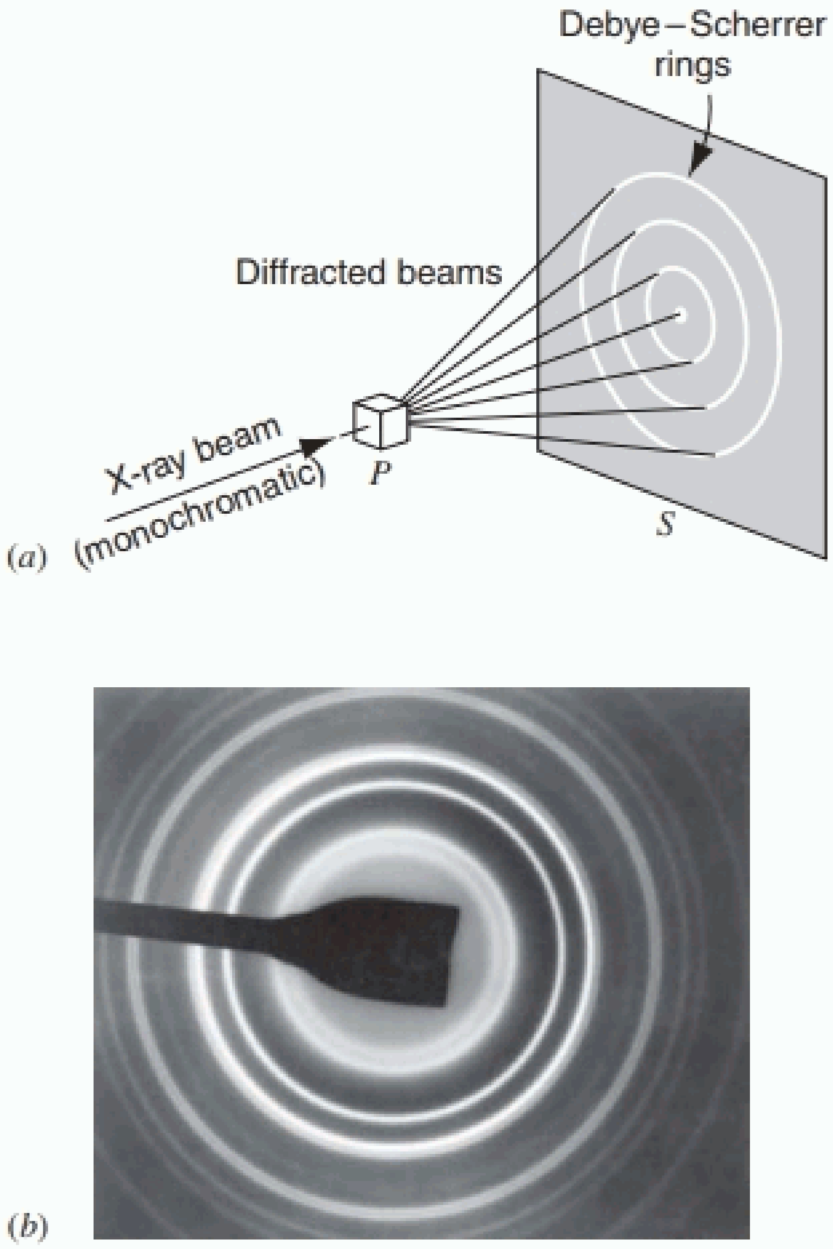


FIGURE 43-24. Question 23.

24. A monochromatic hologram is made with a red He-Ne laser. The hologram film is then illuminated with *green* laser light. Describe the image, if any, that is produced.

EXERCISES

43-1 Multiple Slits

1. A diffraction grating 21.5 mm wide has 6140 rulings. (*a*) Calculate the distance d between adjacent rulings. (*b*) At what angles will maximum-intensity beams occur if the incident radiation has a wavelength of 589 nm?

2. A diffraction grating 2.86 cm wide produces a deviation of 33.2° in the second order with light of wavelength 612 nm. Find the total number of rulings on the grating.

3. With light from a gaseous discharge tube incident normally on a grating with a distance 1.73 μm between adjacent slit centers, a green line appears with sharp maxima at measured transmission angles $\theta = \pm 17.6°$, $37.3°$, $-37.1°$, $65.2°$, and $-65.0°$. Compute the wavelength of the green line that best fits the data.

4. A narrow beam of monochromatic light strikes a grating at normal incidence and produces sharp maxima at the following angles from the normal: 6°40′, 13°30′, 20°20′, 35°40′. No other maxima appear at any angle between 0° and 35°40′. The separation between adjacent ruling centers in the grating is 5040 nm. Find the wavelength of light used.

5. Light of wavelength 600 nm is incident normally on a diffraction grating. Two adjacent principal maxima occur at $\sin\theta = 0.20$ and $\sin\theta = 0.30$. The fourth order is missing. (*a*) What is the separation between adjacent slits? (*b*) What is the smallest possible individual slit width? (*c*) Name all orders actually appearing on the screen with the values derived in (*a*) and (*b*).

6. A diffraction grating is made up of slits of width 310 nm with a 930-nm separation between centers. The grating is illuminated by monochromatic plane waves, $\lambda = 615$ nm, with the angle of incidence zero. (*a*) How many diffraction maxima are there? (*b*) Find the width of the spectral lines observed in first order if the grating has 1120 slits.

43-2 Diffraction Gratings

7. A grating has 315 rulings/mm. For what wavelengths in the visible spectrum can fifth-order diffraction be observed?

8. A diffraction grating has 200 rulings/mm and a principal maximum is noted at $\theta = 28°$. (*a*) What are the possible wavelengths of the incident visible light? (*b*) What colors are they?

9. Given a grating with 400 rulings/mm, how many orders of the entire visible spectrum (400–700 nm) can be produced?

10. Show that, in a grating with alternately transparent and opaque strips of equal width, all the even orders (except $m = 0$) are absent.

11. White light ($400 \text{ nm} < \lambda < 700 \text{ nm}$) is incident on a grating. Show that, no matter what the value of the grating spacing d, the second- and third-order spectra overlap.

12. Assume that light is incident on a grating at an angle ψ as shown in Fig. 43-25. Show that the condition for a diffraction maximum is

$$d(\sin\psi + \sin\theta) = m\lambda \qquad m = 0, 1, 2, \ldots .$$

Only the special case $\psi = 0$ has been treated in this chapter (compare with Eq. 43-1).

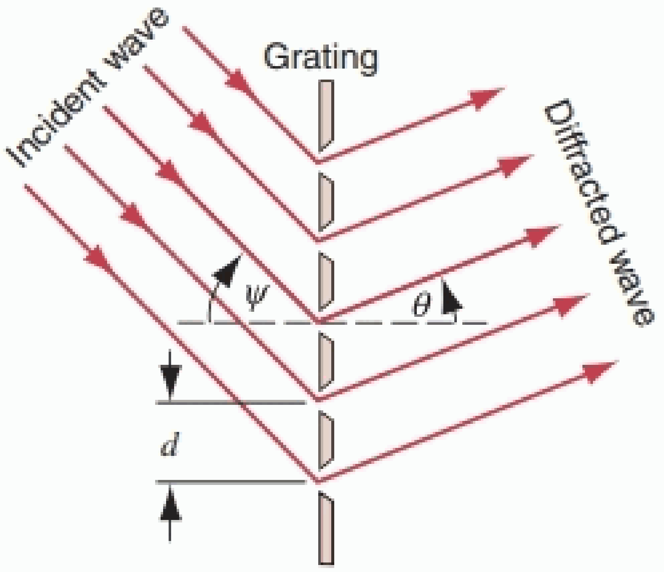


FIGURE 43-25. Exercise 12.

13. A transmission grating with $d = 1.50$ μm is illuminated at various angles of incidence by light of wavelength 600 nm. Plot as a function of angle of incidence (0° to 90°) the angular deviation of the first-order diffracted beam from the incident direction. See Exercise 12.

14. Assume that the limits of the visible spectrum are arbitrarily chosen as 430 and 680 nm. Calculate the number of rulings per mm of a grating that will spread the first-order spectrum through an angular range of 20.0°.

15. A grating has 350 rulings/mm and is illuminated at normal incidence by white light. A spectrum is formed on a screen 30 cm from the grating. If a 10-mm square hole is cut in the screen, with its inner edge 50 mm from the central maximum and parallel to it, what range of wavelengths passes through the hole?

43-3 Dispersion and Resolving Power

16. A grating has 620 rulings/mm and is 5.05 mm wide. (*a*) What is the smallest wavelength interval that can be resolved in the third order at $\lambda = 481$ nm? (*b*) How many higher orders can be seen?

17. The "sodium doublet" in the spectrum of sodium is a pair of lines with wavelengths 589.0 and 589.6 nm. Calculate the minimum number of rulings in a grating needed to resolve this doublet in the second-order spectrum.

18. (*a*) How many rulings must a 4.15-cm-wide diffraction grating have to resolve the wavelengths 415.496 nm and 415.487 nm in the second order? (*b*) At what angle are the maxima found?

19. A source containing a mixture of hydrogen and deuterium atoms emits light containing two closely spaced red colors at $\lambda = 656.3$ nm whose separation is 0.180 nm. Find the minimum number of rulings needed in a diffraction grating that can resolve these lines in the first order.

20. Shows that the dispersion of a grating can be written as

$$D = \lambda^{-1} \tan\theta.$$

21. In a particular grating, the sodium doublet (see Exercise 17) is viewed in third order at 10.2° to the normal and is barely resolved. Find (*a*) the ruling spacing and (*b*) the total width of the grating.

22. Light containing a mixture of two wavelengths, 500 nm and 600 nm, is incident normally on a diffraction grating. It is desired (1) that the first and second principal maxima for each wavelength appear at $\theta \leq 30°$, (2) that the dispersion be as high as possible, and (3) that the third order for 600 nm be a missing order. (*a*) What should be the separation between adjacent slits? (*b*) What is the smallest possible individual slit width? (*c*) Name all orders for 600 nm that actually appear on the screen with the values derived in (*a*) and (*b*).

23. A grating has 40,000 rulings spread over 76 mm. (*a*) What is its expected dispersion D in °/nm for sodium light ($\lambda = 589$ nm) in the first three orders? (*b*) What is its resolving power in these orders?

43-4 X-ray Diffraction

24. X rays of wavelength 0.122 nm are found to reflect in the second order from the face of a lithium fluoride crystal at a Bragg angle of 28.1°. Calculate the distance between adjacent crystal planes.

25. A beam of x rays of wavelength 29.3 pm is incident on a calcite crystal of lattice spacing 0.313 nm. Find the smallest angle between the crystal planes and the beam that will result in constructive reflection of the x rays.

26. Monochromatic, high-energy x rays are incident on a crystal. If first-order reflection is observed at Bragg angle 3.40°, at what angle would second-order reflection be expected?

27. An x-ray beam, containing radiation of two distinct wavelengths, is scattered from a crystal, yielding the intensity spectrum shown in Fig. 43-26. The interplanar spacing of the scattering planes is 0.94 nm. Determine the wavelengths of the x rays present in the beam.

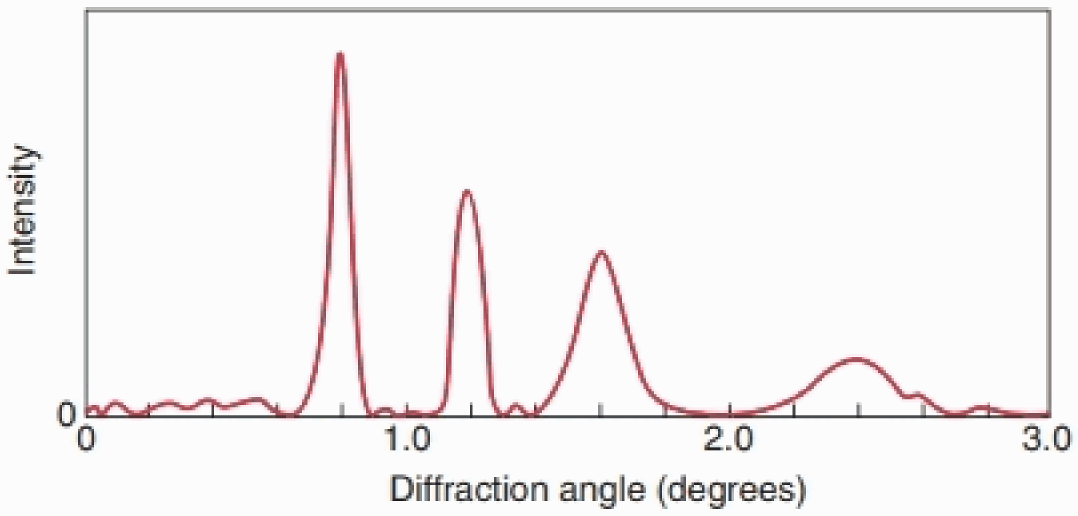


FIGURE 43-26. Exercise 27.

28. In comparing the wavelengths of two monochromatic x-ray lines, it is noted that line A gives a first-order reflection maximum at a glancing angle of 23.2° to the face of a crystal. Line B, known to have a wavelength of 96.7 pm, gives a third-order reflection maximum at an angle of 58.0° from the same face of the same crystal. (*a*) Calculate the interplanar spacing. (*b*) Find the wavelength of line A.

29. Monochromatic x rays are incident on a set of NaCl crystal planes whose interplanar spacing is 39.8 pm. When the beam is rotated 51.3° from the normal, first-order Bragg reflection is observed. Find the wavelength of the x rays.

30. Show that, in Bragg diffraction by a monochromatic beam of x rays, no intense maxima will be obtained if the wavelength of the x rays is greater than twice the largest crystal plane separation. See Question 19.

31. Prove that it is not possible to determine both wavelength of radiation and spacing of Bragg reflecting planes in a crystal by measuring the angles for Bragg reflection in several orders.

32. Assume that the incident x-ray beam in Fig. 43-27 is not monochromatic but contains wavelengths in a band from 95.0 to 139 pm. Will diffracted beams, associated with the planes shown, occur? If so, what wavelengths are diffracted? Assume $d = 275$ pm.

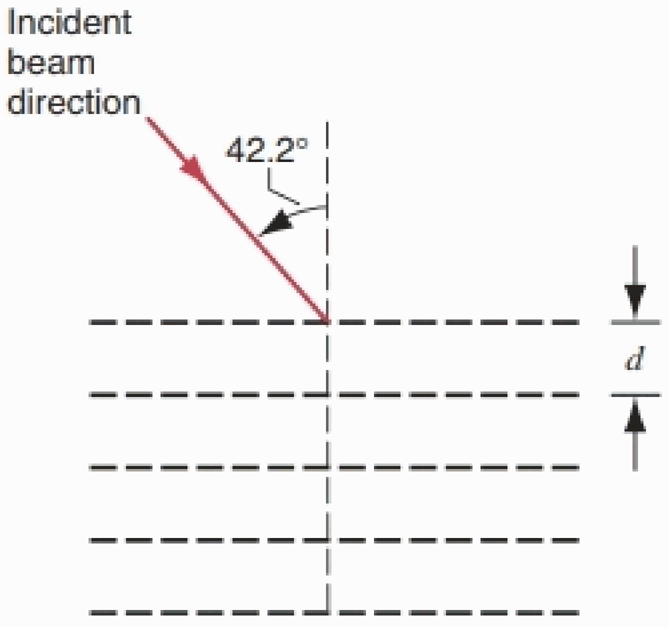


FIGURE 43-27. Exercises 32 and 34.

33. First-order Bragg scattering from a certain crystal occurs at an angle of incidence of 63.8°; see Fig. 43-28. The wavelength of the x rays is 0.261 nm. Assuming that the scattering is from the dashed planes shown, find the unit cell size a_0.

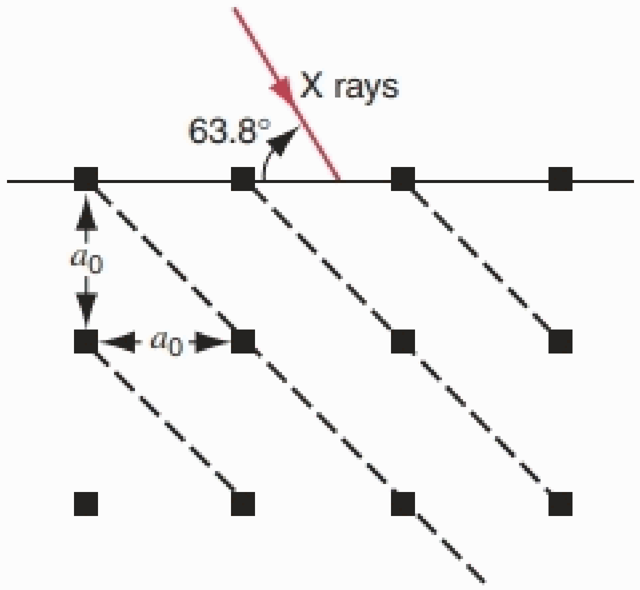


FIGURE 43-28. Exercise 33.

34. Monochromatic x rays ($\lambda = 0.125$ nm) fall on a crystal of sodium chloride, making an angle of 42.2° with the reference line shown in Fig. 43-27. The planes shown are those of Fig. 43-18*a*, for which $d = 0.252$ nm. Through what angles must the crystal be turned to give a diffracted beam associated with the planes shown? Assume that the crystal is turned about an axis that is perpendicular to the plane of the page.

43-5 Holography

PROBLEMS

1. Derive this expression for the intensity pattern for a three-slit "grating":

$$I = \frac{1}{9} I_m (1 + 4 \cos \phi + 4 \cos^2 \phi),$$

where

$$\phi = \frac{2\pi d \sin \theta}{\lambda}.$$

Assume that $a \ll \lambda$ and be guided by the derivation of the corresponding double-slit formula (Eq. 42-17).

2. (*a*) Using the result of Problem 1, show that the half-width of the fringes for a three-slit diffraction pattern, assuming θ small enough so that $\sin \theta \approx \theta$, is

$$\Delta\theta \approx \frac{\lambda}{3.2d}.$$

(*b*) Compare this with the expression derived for the two-slit pattern in Problem 5, Chapter 41 and show that these results support the conclusion that for a fixed slit spacing the interference maxima become sharper as the number of slits is increased.

3. (*a*) Using the result of Problem 1, show that a three-slit "grating" has only one secondary maximum. Find (*b*) its location and (*c*) its relative intensity.

4. A three-slit grating has separation d between adjacent slits. If the middle slit is covered up, will the half-width of the intensity maxima become broader or narrower and by what factor? See Problem 2.

5. A diffraction grating has a large number N of slits, each of width a. Let $I_{\max}$ denote the intensity at some principal maximum, and let I_k denote the intensity of the kth adjacent secondary maximum. (*a*) If $k \ll N$, show from the phasor diagram that, approximately, $I_k/I_{\max} = 1/(k + \frac{1}{2})^2\pi^2$. (Compare this with the single-slit formula.) (*b*) For those secondary maxima that lie roughly midway between two adjacent principal maxima, show that roughly $I_k/I_{\max} = 1/N^2$. (*c*) Consider the central principal maximum and those adjacent secondary maxima for which $k \ll N$. Show that this part of the diffraction pattern quantitatively resembles that for one single slit of width Na.

6. A diffraction grating has a resolving power $R = \lambda/\Delta\lambda = Nm$. (*a*) Show that the corresponding frequency range Δf that can just be resolved is given by $\Delta f = c/Nm\lambda$. (*b*) From Fig. 43-2, show that the "times of flight" of the two extreme rays differ by an amount $\Delta t = (Nd/c)\sin \theta$. (*c*) Show that $(\Delta f)(\Delta t) = 1$, with this relation independent of the various grating parameters. Assume $N \gg 1$.

7. Consider an infinite two-dimensional square lattice as in Fig. 43-16*b*. One interplanar spacing is obviously a_0 itself. (*a*) Calculate the next five smaller interplanar spacings by sketching figures similar to Fig. 43-18*a*. (*b*) Show that the general formula is

$$d = a_0/\sqrt{h^2 + k^2},$$

where h and k are both relatively prime integers that have no common factors other than unity.

8. In Sample Problem 43-5, the $m = 1$ beam, permitted by interference considerations, has zero intensity because of the diffracting properties of the unit cell for this geometry of beams and crystal. Prove this. (Hint: Show that the "reflection" from an atomic plane through the top of a layer of unit cells is canceled by a "reflection" from a plane through the middle of this layer of cells. All odd-order beams prove to have zero intensity.)

CHAPTER 44

POLARIZATION

In Chapter 38, we showed electromagnetic waves traveling such that the electric field vector $\vec{\mathbf{E}}$ and magnetic field vector $\vec{\mathbf{B}}$ are perpendicular to each other and to the direction of propagation of the wave. That is, electromagnetic waves are transverse *waves. This prediction follows from Maxwell's equations.*

In many of the experiments we have described so far, light waves do not reveal their transverse nature. For example, reflection, refraction, interference, and diffraction can occur for longitudinal waves (such as sound) as well as for transverse waves. In this chapter we consider phenomena that not only depend on the transverse nature of electromagnetic radiation, they also depend on the particular orientation in space of the electric and magnetic field vectors.

44-1 POLARIZATION OF ELECTROMAGNETIC WAVES

You may have had the experience of rotating a small portable radio one way or another to improve the reception of the station to which you were listening. Or perhaps you adjusted the position or orientation of an antenna attached to your stereo system to increase the strength of the signal. If you have a TV antenna on your roof, you must align it in the proper orientation to receive the signals that are broadcast from the station. These adjustments are necessary because some types of antennas respond to the electric field $\vec{\mathbf{E}}$ of an electromagnetic wave, and a signal can be received only if the electric field of the wave can force electrons to flow along wires to make a current. The orientation of the antenna must be chosen to match the orientation of the $\vec{\mathbf{E}}$ field of the wave as it is produced by the broadcast antenna. For example, in the United States, television signals may be broadcast so that the $\vec{\mathbf{E}}$ field oscillates in a horizontal plane, so the plane of the rooftop antenna must also be horizontal (Fig. 44-1). In some other countries, TV signals may be broadcast with the $\vec{\mathbf{E}}$ field oscillating in a vertical plane, and so a different orientation of the rooftop antenna would be required to receive the signal.

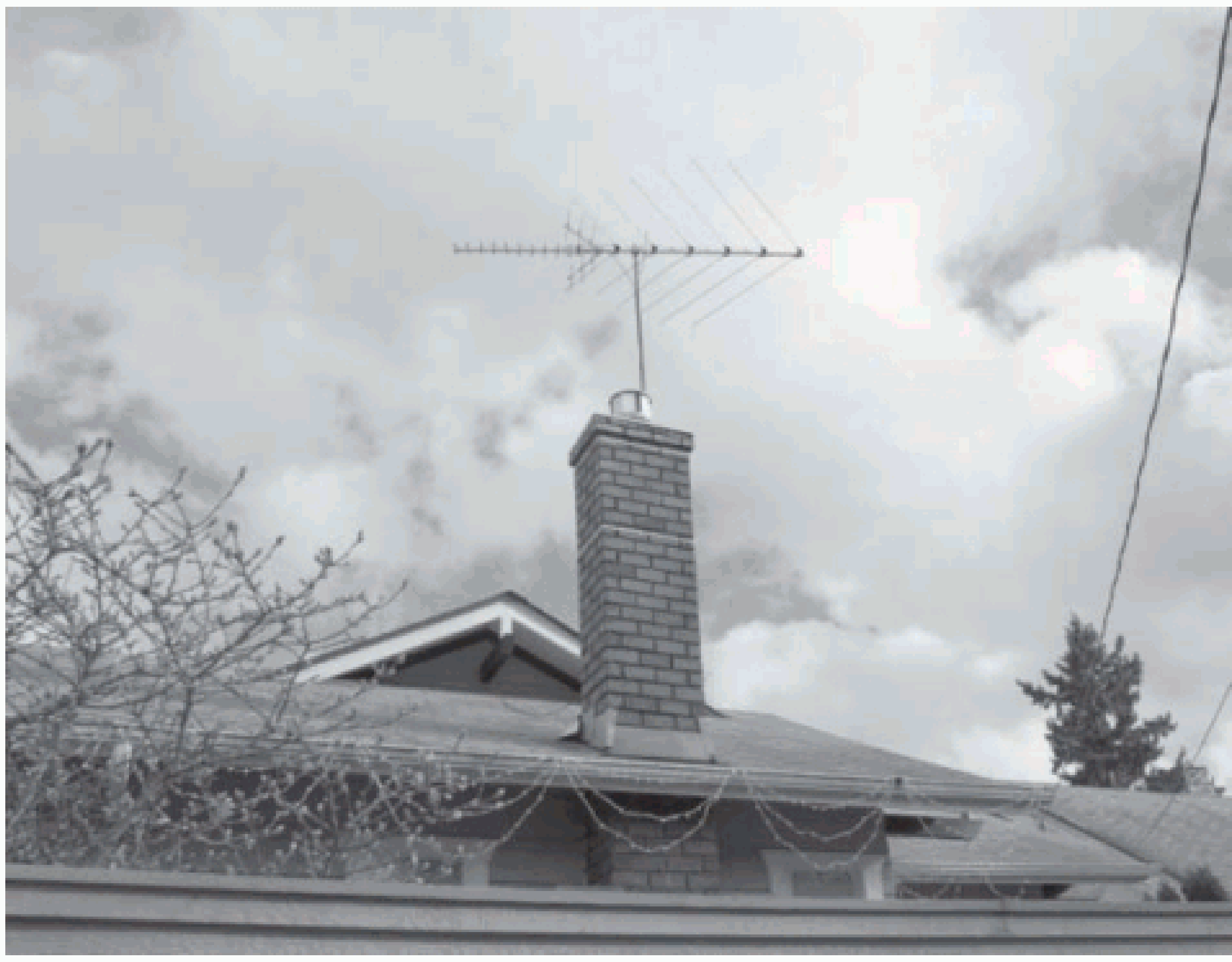


FIGURE 44-1. Rooftop TV antennas in the United States.

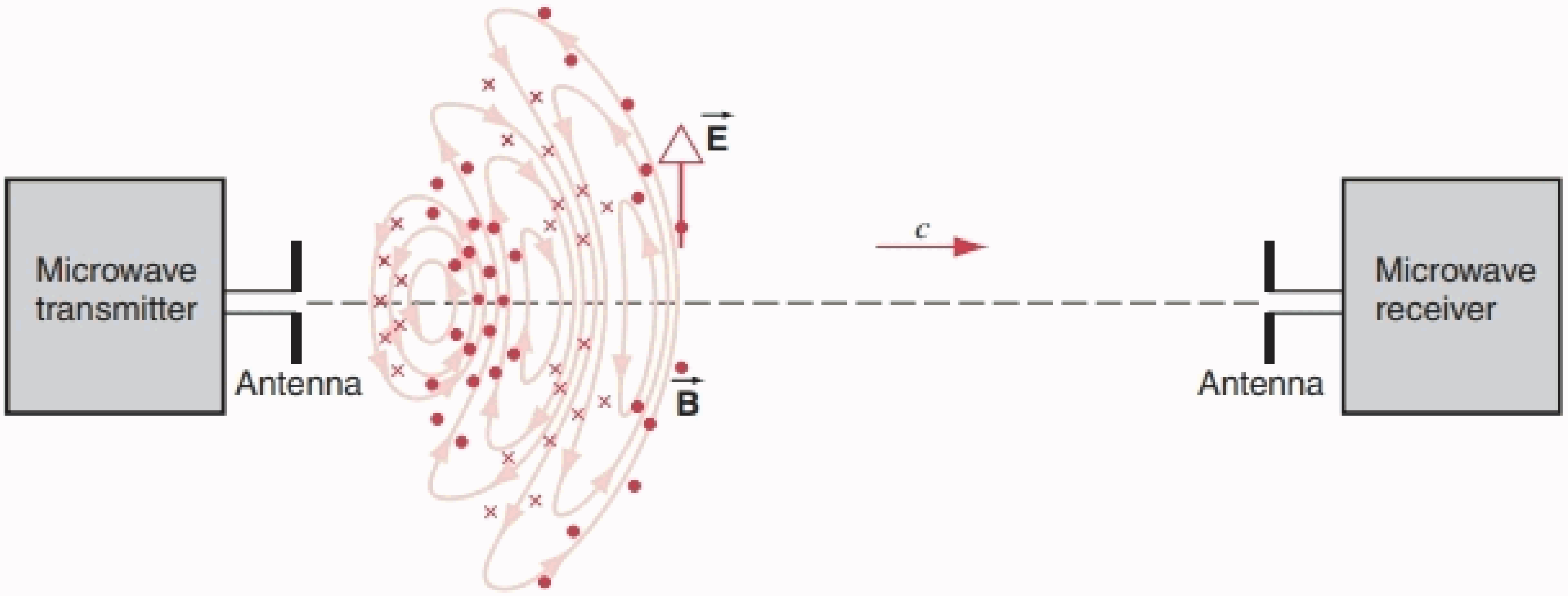


FIGURE 44-2. The electromagnetic wave generated by the transmitter is polarized in the plane of the page, with its $\vec{\mathbf{E}}$ vector parallel to the axis of the transmitting antenna. The receiving antenna can detect this wave with maximum effectiveness if its antenna also lies in the plane and parallel to $\vec{\mathbf{E}}$. If the receiving antenna were rotated through 90° about the direction of propagation, no signal would be detected.

Figure 44-2 shows an experimental arrangement that demonstrates this effect. A microwave transmitter on the left is connected to a dipole antenna. Charges surging up and down in the antenna produce an electromagnetic wave whose $\vec{\mathbf{E}}$ vector is (at large distances from the dipole) parallel to its axis. When this wave is incident on the antenna of the microwave receiver at the right, the $\vec{\mathbf{E}}$ vector of the wave causes charges to move up and down in the antenna. These moving charges produce a signal in the receiver.

If the transmitter were rotated by 90° about the direction of propagation of the wave, the signal in the receiver would drop to zero. In this case, the $\vec{\mathbf{E}}$ vector of the wave would be at right angles to the axis of the receiving antenna; the wave would produce no movement of charge along the antenna and thus no signal in the receiver. A similar result would be obtained if the receiver were rotated instead of the transmitter.

Figure 44-3 represents a plane electromagnetic wave like the one that would be observed far from the transmitting antenna in Fig. 44-2. As is always the case, the $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ vectors are perpendicular to one another and to the direction of propagation of the wave, which is the basic picture of a transverse wave. By convention, we define the *direction of polarization* of the wave to be the direction of the $\vec{\mathbf{E}}$ vector (the y direction in Fig. 44-3). The plane determined by the $\vec{\mathbf{E}}$ vector and the direction of propagation of the wave (the xy plane in Fig. 44-3) is called the *plane of polarization* of the wave. Note that specifying two directions of an electromagnetic wave (the direction of propagation and the direction of $\vec{\mathbf{E}}$) completely specifies the wave, because the direction of $\vec{\mathbf{B}}$ is fixed by these two directions.*

The wave illustrated in Fig. 44-3 is said to be *linearly polarized* (also called plane polarized). This means that the $\vec{\mathbf{E}}$ field remains in a fixed direction (the y direction in Fig. 44-3) as the wave propagates. As in the experiment shown in Fig. 44-2, linearly polarized electromagnetic waves, such as those that are used for radio and TV transmission, can be produced by orienting the axis of a dipole transmitting antenna in a certain direction. The motions of the electrons in the antenna are *coherent;* they act in unison to transmit a polarized electromagnetic wave. Figure 44-4*a* shows a view of the linearly polarized wave of Fig. 44-3 as it might be seen if it were traveling out of the paper toward you. The electric field oscillates in a single direction (in this case, up and down along the y axis).

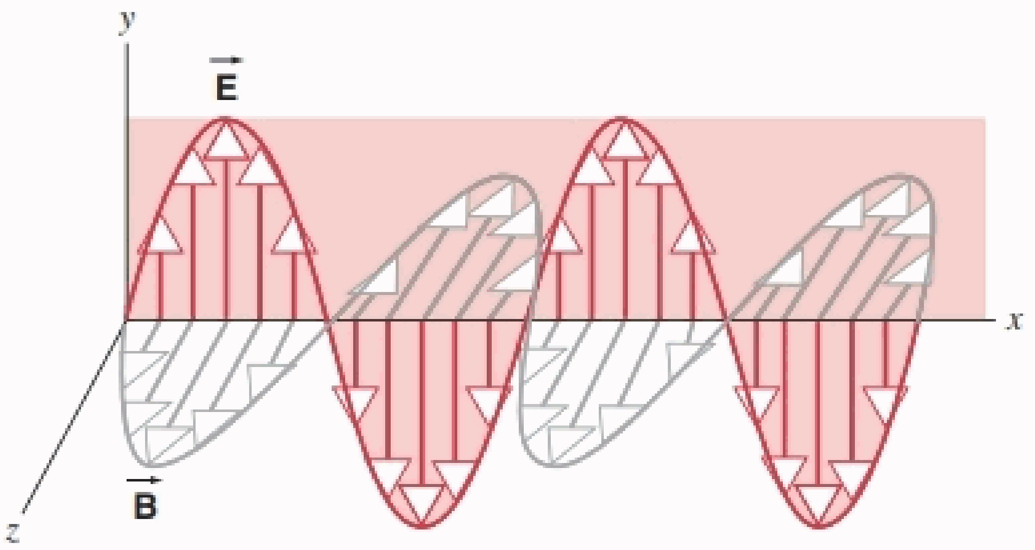


FIGURE 44-3. An instantaneous snapshot of a traveling electromagnetic wave showing the $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ vectors. The wave is linearly polarized, in this case in the y direction. The *plane of polarization* is defined to be the plane containing the $\vec{\mathbf{E}}$ vector and the direction of propagation; in this case, the plane of polarization is the xy plane.

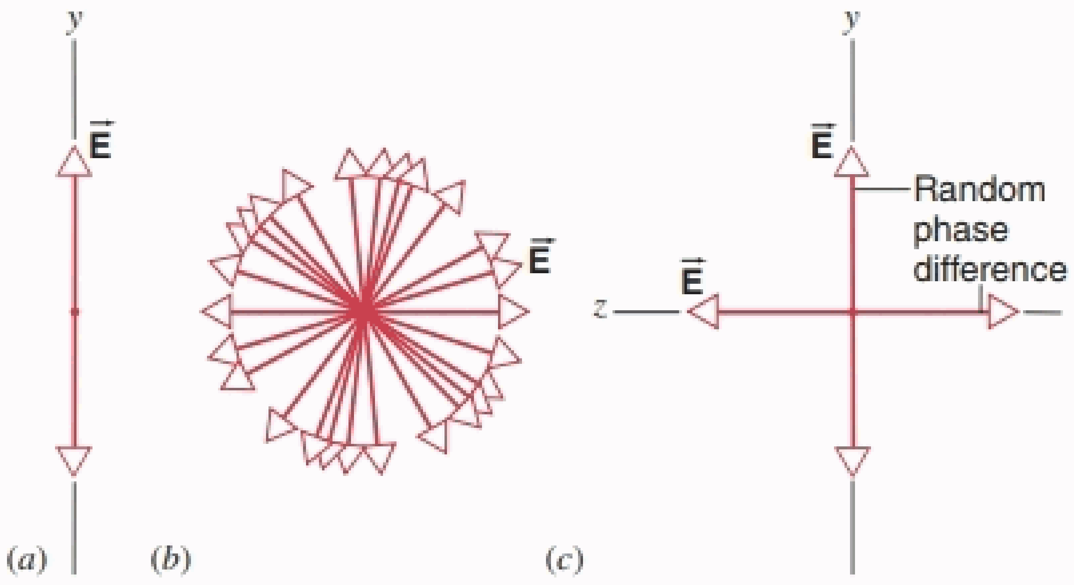


FIGURE 44-4. (*a*) A linearly polarized wave, such as that of Fig. 44-3, viewed from along the direction of propagation. The wave is moving out of the plane of the page. Only the direction of the $\vec{\mathbf{E}}$ vector is shown. (*b*) An unpolarized wave, which can be considered to be a random superposition of many polarized waves. (*c*) An equivalent way of showing the unpolarized wave, as two waves linearly polarized at right angles to one another and with a random phase difference between them. The orientation of the y and z axes about the direction of propagation is completely arbitrary.

---

* Recall the Poynting vector, $\vec{\mathbf{S}} = (\vec{\mathbf{E}} \times \vec{\mathbf{B}})/\mu_0$, discussed in Section 38-6, where $\vec{\mathbf{S}}$ is in the direction of propagation of the wave. Given $\vec{\mathbf{S}}$ and $\vec{\mathbf{E}}$, we can find the magnitude and direction of $\vec{\mathbf{B}}$.

In other sources of electromagnetic waves, such as the Sun or an incandescent lightbulb, the atoms behave independently rather than cooperatively. Taken together, the emissions from these sources do not show a single plane of polarization; instead, the radiation is composed of many components whose polarizations are randomly distributed around the direction of propagation, such as is represented in Fig. 44-4*b*. This light is transverse but *unpolarized;* that is, there is no preferred plane of polarization.

Figure 44-4*c* shows another way to represent an unpolarized wave. The random $\vec{\mathbf{E}}$ vectors are represented by their net components on any two perpendicular axes (in this case, y and z). For unpolarized waves, the components have equal amplitudes, and the phase difference between them varies randomly with time.

44-2 POLARIZING SHEETS

Figure 44-5 shows unpolarized light falling on a sheet of commercial polarizing material called *Polaroid.** There exists in the sheet a certain characteristic polarizing direction, shown by the parallel lines. The sheet transmits only those wavetrain components whose electric field vectors vibrate parallel to this direction and absorbs those that vibrate at right angles to this direction. The light emerging from the sheet is linearly polarized. The polarizing direction of the sheet is established during the manufacturing process by embedding certain long-chain molecules in a flexible plastic sheet and then stretching the sheet so that the molecules are aligned parallel to each other. Radiation with its $\vec{\mathbf{E}}$ vector parallel to the long molecules is strongly absorbed, while radiation with its $\vec{\mathbf{E}}$ vector perpendicular to that direction passes through the sheet.

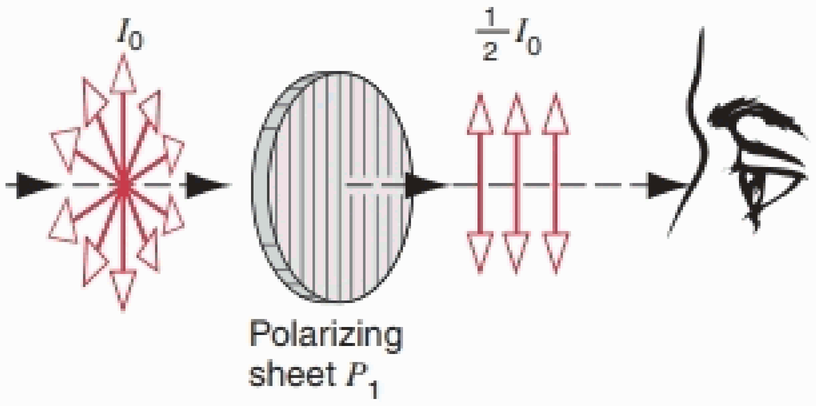


FIGURE 44-5. Unpolarized light is linearly polarized (and reduced in intensity by half) after passing through a single polarizing sheet. The parallel lines, which are not actually visible on the sheet, suggest its polarization direction.

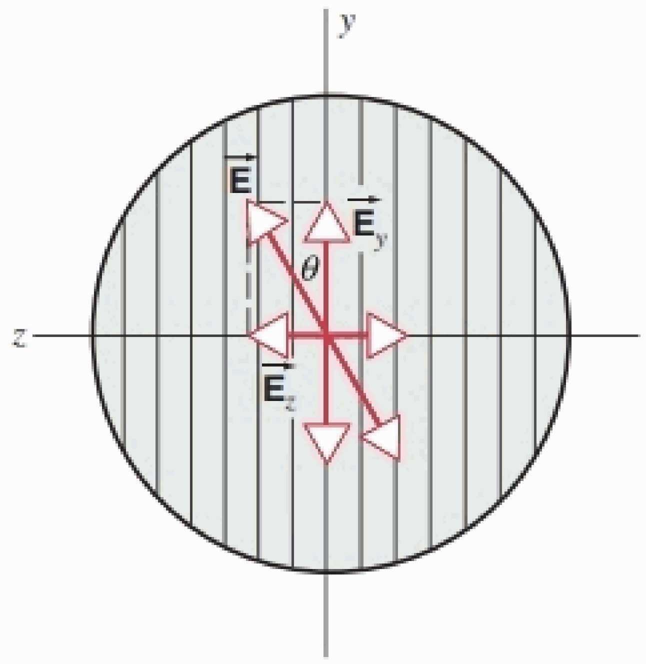


FIGURE 44-6. Another view of the action of a polarizing sheet. A linearly polarized wave (perhaps one of those shown in Fig. 44-4*b*) oriented in a random direction θ falls on the sheet. The y component of $\vec{\mathbf{E}}$ is transmitted, and the z component is absorbed.

In Fig. 44-6, the polarizing sheet or *polarizer* lies in the plane of the page, and the direction of propagation is out of the page. The vector $\vec{\mathbf{E}}$ shows the plane of vibration of a randomly selected wavetrain falling on the sheet. Two vector components, $\vec{\mathbf{E}}_z$ (of magnitude $E \sin \theta$) and $\vec{\mathbf{E}}_y$ (of magnitude $E \cos \theta$), can replace $\vec{\mathbf{E}}$, one parallel to the polarizing direction and one at right angles to it. Only the component $\vec{\mathbf{E}}_y$ is transmitted; the component $\vec{\mathbf{E}}_z$ is absorbed within the sheet.

When unpolarized light is incident on an ideal polarizing sheet, the intensity of the polarized light transmitted through the sheet is half the incident intensity, no matter what the orientation of the sheet is. We can see this from the representation of the incident unpolarized light given in Fig. 44-4*c*, in which each of the components has, on the average, half the intensity of the incident light. Because the orientation of the axes in Fig. 44-4*c* is arbitrary, we are free to choose one of them to be along the direction of transmission of the polarizing sheet on which it is incident. Since this component of the light would be completely transmitted and the other completely absorbed, the sheet transmits 50% of the incident light. We can reach the same conclusion from Fig. 44-6, in which a wave polarized in an arbitrary direction is incident on a polarizing sheet. The component E_y $(= E \cos \theta)$ is transmitted, so the transmitted intensity is proportional to $E_y^2 = E^2 \cos^2 \theta$. If the incident light is unpolarized, we find the total transmitted intensity by averaging this expression over all possible orientations of the plane of polarization of the incident light—that is, over all possible values of θ. The average value of $\cos^2 \theta$ is $\frac{1}{2}$, so we again conclude that half the incident light is transmitted. Because of reflection and partial absorption of the light along the polarizing direction, real polarizing sheets may transmit only 40% of the incident intensity. In our discussions, we assume ideal polarizers.

Let us place a second polarizing sheet P_2 (usually called, when used in this way, an *analyzer*) as in Fig. 44-7. If P_2 is rotated about the direction of propagation, there are two positions, 180° apart, at which the transmitted light intensity falls to zero; these are the positions in which the polarizing directions of P_1 and P_2 are at right angles.

If the amplitude of the linearly polarized light incident on P_2 is E_{m}, the amplitude of the light that emerges is $E_{\mathrm{m}} \cos \theta$, where θ is the angle between the polarizing

---

* There are other ways of producing polarized light without using this well-known commercial product. We mention some of them later. Also see "The Amateur Scientist," by Jearl Walker, *Scientific American,* December 1977, p. 172, for ways of making polarizing sheets and quarter-wave and half-wave plates and for various experiments that can be done with them.

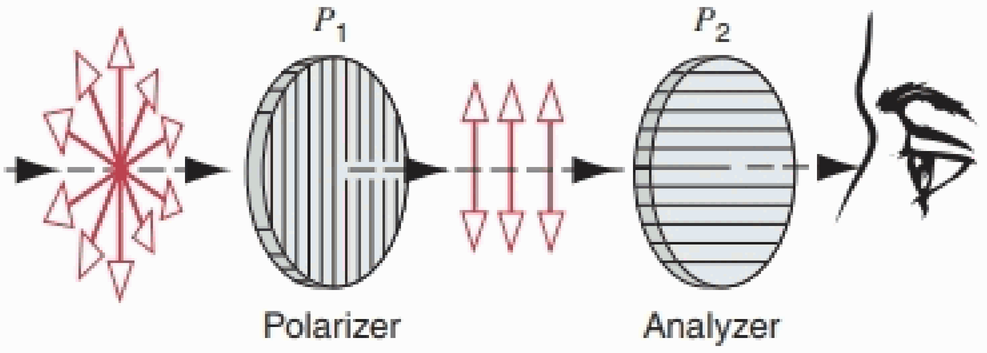


FIGURE 44-7. Unpolarized light is not transmitted by two polarizing sheets whose polarizing directions are perpendicular to one another.

directions of P_1 and P_2. Recalling that the intensity of the light beam is proportional to the square of the amplitude, we see that the transmitted intensity I varies with θ according to

$$I = I_m \cos^2 \theta, \qquad (44\text{-}1)$$

in which I_m, the maximum value of the transmitted intensity, occurs when the polarizing directions of P_1 and P_2 are parallel—that is, when $\theta = 0$ or 180°. Figure 44-8*a*, in which two overlapping polarizing sheets are in the parallel position ($\theta = 0$ or 180° in Eq. 44-1), shows that the intensity of the light transmitted through the region of overlap has its maximum value. In Fig. 44-8*b*, one of the sheets has been rotated through 90° so that θ in Eq. 44-1 has the value 90° or 270°; the intensity of the light transmitted through the region of overlap is now a minimum. Equation 44-1, called the *law of Malus,* was discovered by Etienne Louis Malus (1775–1812) experimentally in 1809, using different polarizing techniques (see Section 44-3).

Historically, polarization studies were made (for example, by Malus and by Thomas Young, who, in addition to demonstrating the wave nature of light by double-slit interference, also demonstrated its transverse nature through polarization experiments) to investigate the nature of light. Today we reverse the procedure and deduce something about the nature of an object from the polarization state of the light emitted by, or scattered from, that object. It has been possible to deduce, from studies of the polarization of light reflected from them, that the grains of cosmic dust present in our galaxy have been oriented in the weak galactic magnetic field (about 10^{-8} T) so that their long dimension is parallel to this field. Polarization studies have suggested that Saturn's rings consist of ice crystals. The size and shape of virus particles can be determined by the polarization of ultraviolet light scattered from them. Information about the structure of atoms and nuclei is obtained from polarization studies of their emitted radiations in all parts of the electromagnetic spectrum. Thus we have a useful research technique for structures ranging in size from a galaxy (10^{+22} m) to a nucleus (10^{-14} m).

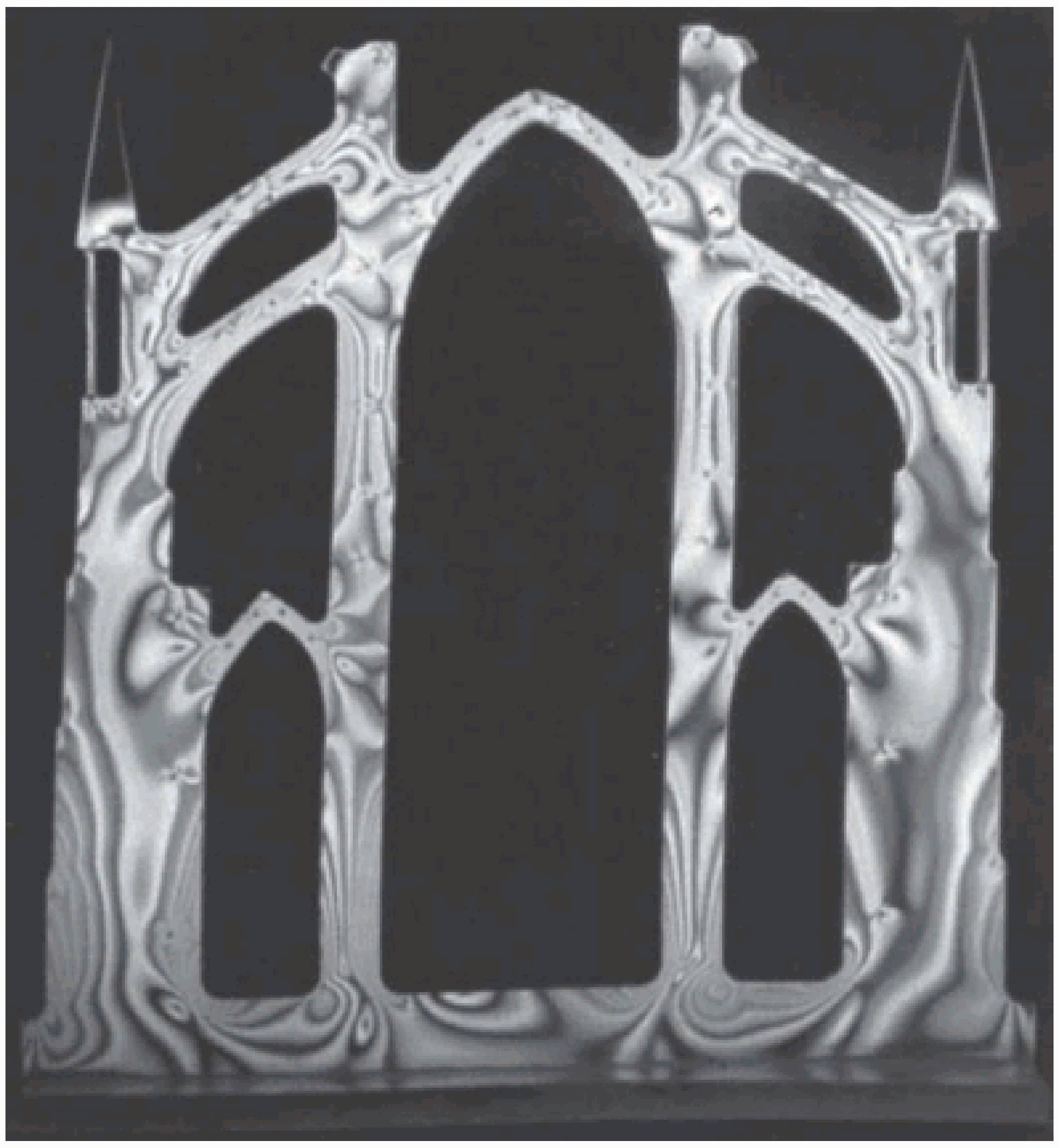


FIGURE 44-9. A piece of plastic is viewed between crossed polarizing sheets. The light and dark patterns show regions of stress in the structure, which is a model of the construction details of a gothic cathedral.

Polarized light also has many applications in industry and in engineering. Figure 44-9 shows a piece of plastic that has been stressed and placed between polarizing sheets. The stress pattern is revealed, allowing engineers to refine their designs to reduce stress at critical locations in the struc-

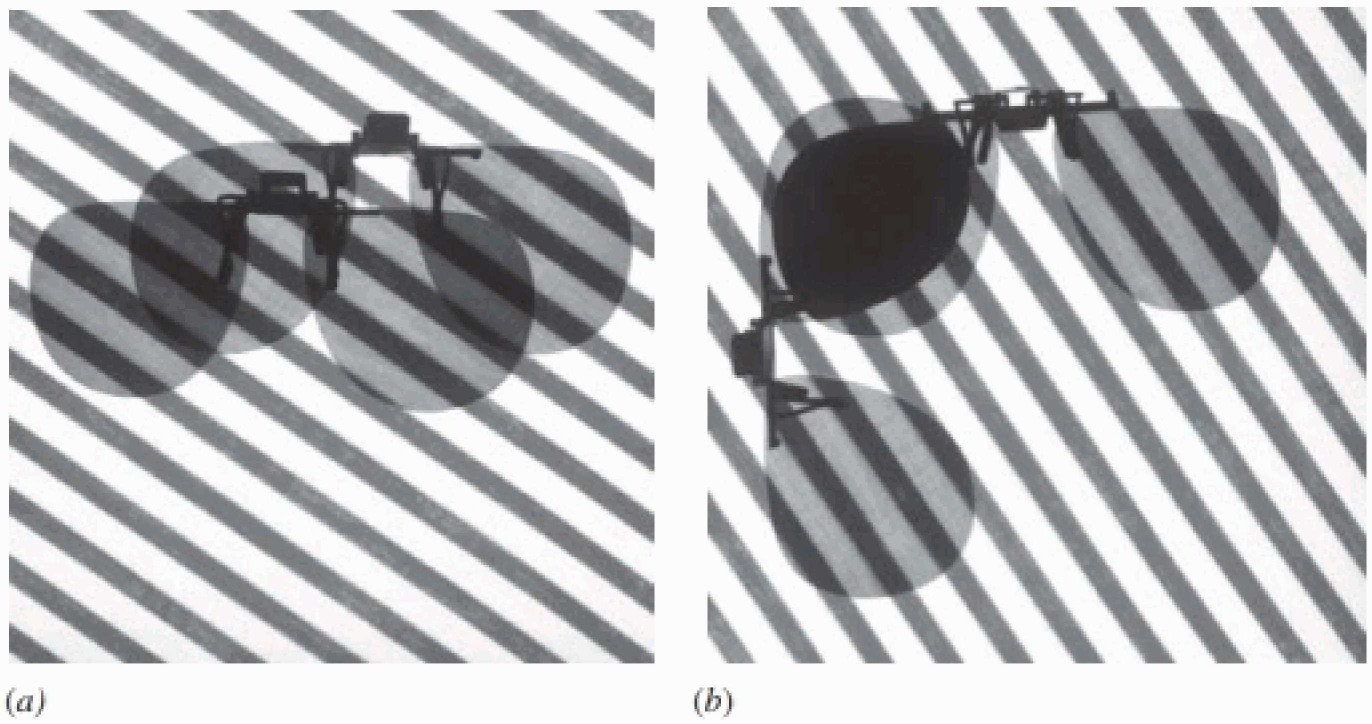


FIGURE 44-8. Two sheets of polarizing material (in this case, polarizing sunglasses) are placed over a pattern of lines. In (*a*) the polarization directions of the two sheets are parallel, so that light passes through; in (*b*) the polarization directions are perpendicular, so that no light passes through the two overlapping lenses.

TEMP
17.8°C
AM
10:58
MONTH
DATE
DAY
10/ 4 WED
CLOCK
DOWN
UP
MODE
PULSE
TOTAL RESET

FIGURE 44-10. A personal digital assistant (PDA) with a liquid crystal display.

ture.* Figure 44-10 shows a common liquid crystal display, which uses polarized light to form letters and numbers, such as on watches and calculator displays. The liquid crystal is a material with stretched molecules like polarizing sheets; however, the long direction can be made to follow an applied electric field. The liquid crystal is arranged so that it normally transmits light through the polarizer and analyzer. When the electric field (from a battery) is applied to certain regions, the molecules line up in such a way that no light is transmitted through those regions, which form the dark patterns of the display.

SAMPLE PROBLEM 44-1. Two polarizing sheets have their polarizing directions parallel so that the intensity I_m of the transmitted light is a maximum. Through what angle must either sheet be turned if the intensity is to drop by one-half?

Solution From Eq. 44-1, since $I = \frac{1}{2}I_m$, we have

$$\tfrac{1}{2}I_m = I_m \cos^2 \theta$$

or

$$\theta = \cos^{-1}\left(\pm\frac{1}{\sqrt{2}}\right) = \pm 45°, \pm 135°.$$

The same effect is obtained no matter which sheet is rotated or in which direction.

44-3 POLARIZATION BY REFLECTION

Malus discovered in 1809 that light can be partially or completely polarized by reflection. Anyone who has watched the Sun's reflection in water, while wearing a pair of sunglasses made of polarizing material, has probably noticed the effect. It is necessary only to tilt the head from side to side, thus rotating the polarizing sheets, to observe that the intensity of the reflected sunlight passes through a minimum.

Figure 44-11 shows an unpolarized beam falling on a glass surface. The $\vec{\mathbf{E}}$ vectors are resolved into two components (as in Fig. 44-4*c*), one perpendicular to the plane of incidence (the plane of Fig. 44-11) and one parallel to this plane. On the average, for completely unpolarized incident light, these two components are of equal amplitude.

For glass or other dielectric materials, there is a particular angle of incidence, called the *polarizing angle* θ_p (also known as *Brewster's angle*), at which the reflection coefficient for the polarization component in the plane of Fig. 44-11 is zero. This means that the beam reflected from the glass, although of low intensity, is linearly polarized, with its plane of polarization perpendicular to the plane of incidence. This polarization of the reflected beam can easily be verified by analyzing it with a polarizing sheet.

When light is incident at the polarizing angle, the component with polarization parallel to the plane of incidence is entirely refracted, while the perpendicular component is partially reflected and partially refracted. Thus the refracted beam,

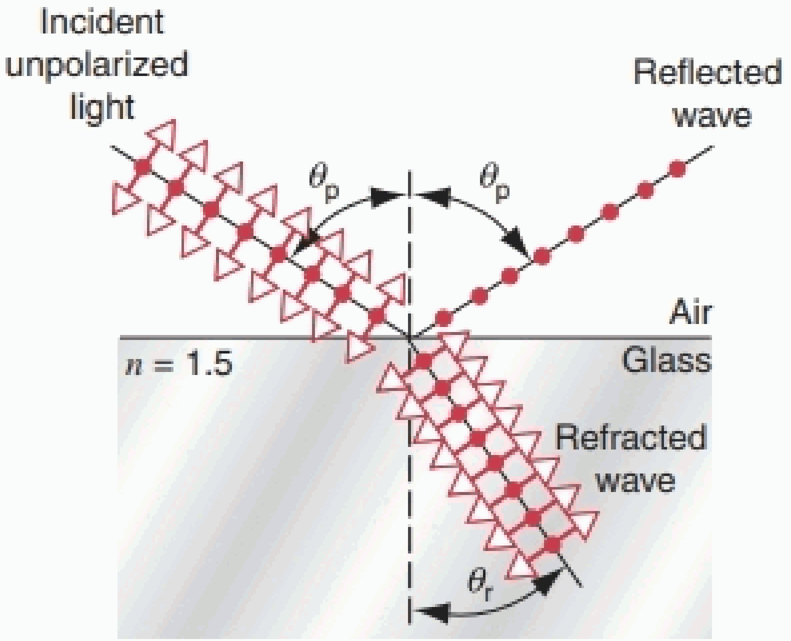


FIGURE 44-11. For a particular angle of incidence θ_p, the reflected light is completely polarized. The refracted light is partially polarized. The dots indicate polarization components perpendicular to the plane of the page, and the double arrows indicate polarization components parallel to the plane of the page.

---

* For examples of how such models are used to study classical architecture, see "The Architecture of Christopher Wren," by Harold Dorn and Robert Mark, *Scientific American,* July 1981, p. 160, and "Gothic Structural Experimentation," by Robert Mark and William W. Clark, *Scientific American,* November 1984, p. 176.

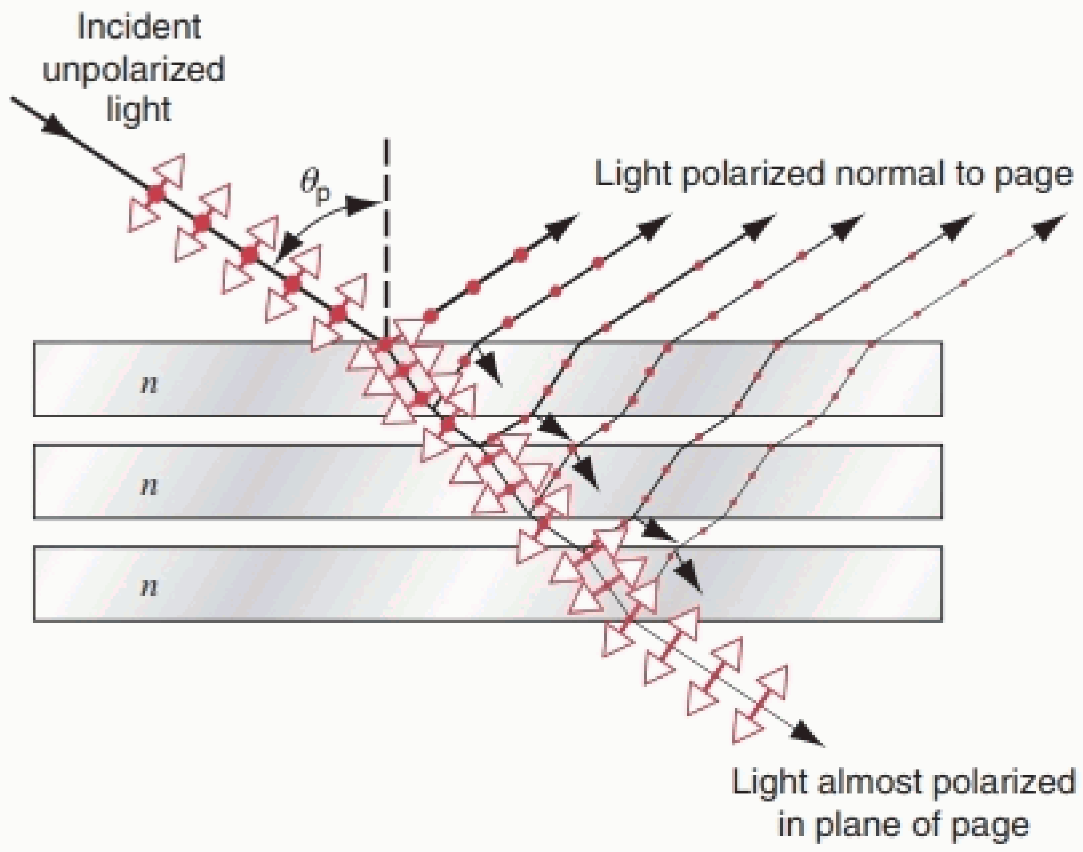

Incident unpolarized light
θ_p
Light polarized normal to page
n
n
n
Light almost polarized in plane of page

FIGURE 44-12. Polarization of light by a stack of glass plates. Unpolarized light is incident at the angle θ_p. All reflected waves are polarized perpendicular to the plane of the page. After passing through several layers, the transmitted wave no longer contains any appreciable component polarized perpendicular to the page.

which is of high intensity, is partially polarized. If this refracted beam passed out of the glass into the air and were then incident on a second glass surface (again at angle θ_p), the perpendicular component would be reflected, and the refracted beam would have a slightly greater polarization. By using a stack of glass plates, we obtain reflections from successive surfaces, and we can increase the intensity of the emerging reflected beam (see Fig. 44-12). The perpendicular components are progressively removed from the transmitted beam, making it more completely polarized in the plane of Fig. 44-12.

At the polarizing angle it is found experimentally that the reflected and the refracted beams are at right angles, or (Fig. 44-11)

$$\theta_p + \theta_r = 90^\circ.$$

From Snell's law,

$$n_1 \sin \theta_p = n_2 \sin \theta_r.$$

Combining these equations leads to

$$n_1 \sin \theta_p = n_2 \sin (90^\circ - \theta_p) = n_2 \cos \theta_p$$

or

$$\tan \theta_p = \frac{n_2}{n_1}, \qquad (44\text{-}2)$$

where the incident beam is in medium 1 and the refracted beam in medium 2. If medium 1 is air ($n_1 = 1$), this becomes

$$\tan \theta_p = n, \qquad (44\text{-}3)$$

where n is the index of refraction of the medium on which the light is incident. Equation 44-2 is known as *Brewster's law* after Sir David Brewster (1781–1868), who deduced it empirically in 1812. It is possible to prove this law rigorously from Maxwell's equations (see also Question 13).

SAMPLE PROBLEM 44-2. We wish to use a plate of glass ($n = 1.50$) in air as a polarizer. Find the polarizing angle and the angle of refraction.

Solution From Eq. 44-3

$$\theta_p = \tan^{-1} 1.50 = 56.3^\circ.$$

The angle of refraction follows from Snell's law:

$$\sin \theta_p = n \sin \theta_r,$$

or

$$\sin \theta_r = \frac{\sin 56.3^\circ}{1.50} = 0.555 \quad \text{or} \quad \theta_r = 33.7^\circ.$$

44-4 DOUBLE REFRACTION

In earlier chapters we assumed that the speed of light, and thus the index of refraction, is independent of the direction of propagation in the medium and of the state of polarization of the light. Liquids, amorphous solids such as glass, and crystalline solids having cubic symmetry normally show this behavior and are said to be *optically isotropic.* Many other crystalline solids are optically *anisotropic* (that is, not isotropic).* Optical anisotropy is responsible for the stress pattern illustrated in Fig. 44-9, although in this case the material is not crystalline.

Figure 44-13, in which a polished crystal of calcite ($CaCO_3$) is laid over a printed pattern, shows the optical

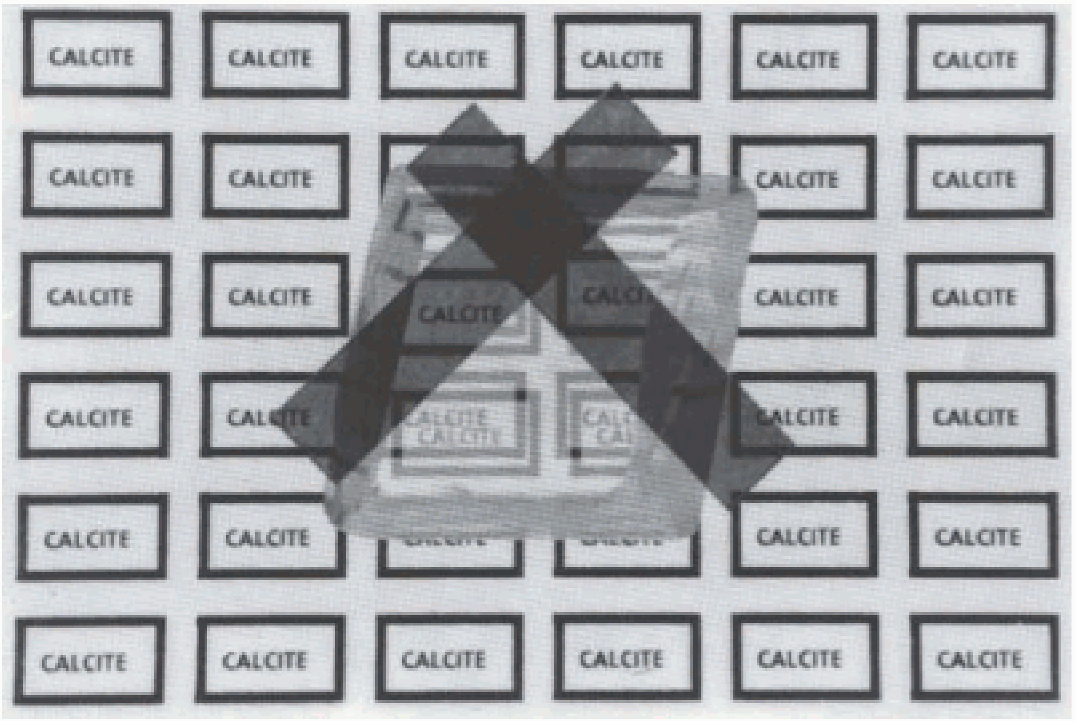


FIGURE 44-13. A view through a birefringent crystal, showing the two images that result from the two different indices of refraction. The double images can be seen where there is no strip of polarizing material. The polarization axis of each strip is parallel to its long direction. Note that the two images have perpendicular polarizations.

---

* Solids may be anisotropic in many properties: mechanical (mica cleaves readily in one plane only), electric (a cube of crystalline graphite does not have the same electric resistance between all pairs of opposite faces), magnetic (a cube of crystalline nickel magnetizes more readily in certain directions than in others), and so forth.

anisotropy of this material; *the image appears double.* Furthermore, the two images show perpendicular polarizations, as indicated in Fig. 44-14, which shows a beam of unpolarized light falling on a calcite crystal at right angles to one of its faces. The single beam splits into two at the crystal surface. The "double-bending" of a beam transmitted through calcite, exhibited in Figs. 44-13 and 44-14, is called *double refraction* or *birefringence.* This phenomenon was studied by Huygens, who described it in his *Treatise on Light* published in 1678.

If the two emerging beams in Fig. 44-14 are analyzed with a polarizing sheet, they are found to be linearly polarized with their planes of vibration at right angles to each other. Figure 44-13 shows that each of the two crossed polarizers transmits only one of the two images (but not the other).

Some doubly refracting materials are strongly absorbing for one polarization component, whereas the other passes through with little absorption. Such materials are called *dichroic.* Polarizing sheets are examples of dichroic material.

If experiments are carried out at various angles of incidence, one of the beams in Fig. 44-14 (represented by the *ordinary ray,* or *o*-ray) is found to obey Snell's law of refraction at the crystal surface, just like a ray passing from one isotropic medium into another. The second beam (represented by the *extraordinary ray,* or *e*-ray) does not obey Snell's law. In Fig. 44-14, for example, the angle of incidence for the incident light is zero but the angle of refraction of the *e*-ray, contrary to the prediction of Snell's law, is nonzero. In general, the *e*-ray does not even lie in the plane of incidence.

This difference between the waves represented by the *o*- and *e*-rays with respect to Snell's law can be explained in these terms:

1. The *o*-wave travels in the crystal with the same speed v_o in all directions. In other words, the crystal has, for this wave, a single index of refraction n_o, just like an isotropic solid.

2. The *e*-wave travels in the crystal with a speed that varies with direction from v_o to v_e. In other words, the index of refraction, defined as c/v, varies with direction from n_o to n_e.

TABLE 44-1 Principal Indices of Refraction of Several Doubly Refracting Crystals[a]

Crystal	*Formula*	n_o	n_e	$n_e - n_o$
Ice	H_2O	1.309	1.313	+0.004
Quartz	SiO_2	1.544	1.553	+0.009
Wurzite	ZnS	2.356	2.378	+0.022
Calcite	$CaCO_3$	1.658	1.486	−0.172
Dolomite	$CaO \cdot MgO \cdot 2CO_2$	1.681	1.500	−0.181
Siderite	$FeO \cdot CO_2$	1.875	1.635	−0.240

[a] For sodium light, $\lambda = 589$ nm.

The quantities n_o and n_e are called the *principal indices of refraction* for the crystal. Problem 7 suggests how to measure them. Table 44-1 shows these indices for six doubly refracting crystals. For three of them the *e*-wave is slower; for the other three it is faster than the *o*-wave. Some doubly refracting crystals (such as mica and topaz) are more complex optically than calcite and require *three* principal indices of refraction for a complete description of their optical properties. Crystals whose basic structure is cubic (such as NaCl; see Fig. 43-15) are optically isotropic, requiring only *one* index of refraction.

The behavior for the speeds of the two waves traveling in calcite is summarized by Fig. 44-15, which shows two Huygens wave surfaces spreading out from an imaginary point light source *S* imbedded in the crystal. The characteristic direction in the crystal in which $v_o = v_e$ is called the *optic axis.* The optic axis is a property of the crystal itself and is independent of the polarization or direction of propagation of the light. The speeds are most different in a direction perpendicular to the optic axis; in that direction the *o*-wave and *e*-wave surfaces have their greatest separation.

The *o*-wave surface in Fig. 44-15 is a sphere, because the medium is isotropic for *o*-waves. The *e*-wave surface cannot be spherical, because the speed of the *e*-wave varies with direction relative to the optic axis. The *e*-wave surface is an ellipsoid of revolution about the optic axis.

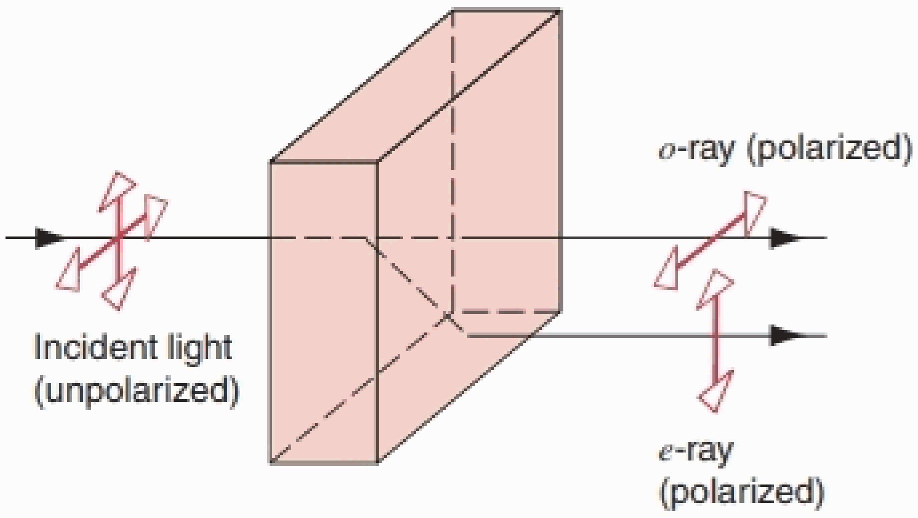


FIGURE 44-14. Unpolarized light falling on a birefringent material (such as a calcite crystal) splits into two components, the *o*-ray (which follows Snell's law of refraction) and the *e*-ray (which does not follow Snell's law). The two refracted rays have perpendicular polarizations, as shown.

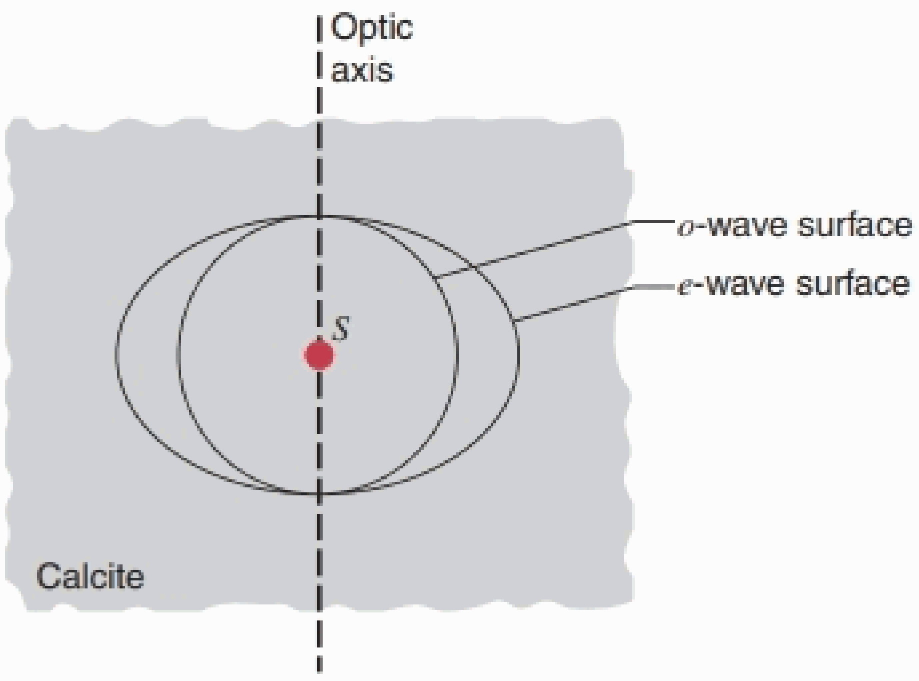


FIGURE 44-15. Huygens wave surfaces produced by a point source *S* imbedded in calcite.

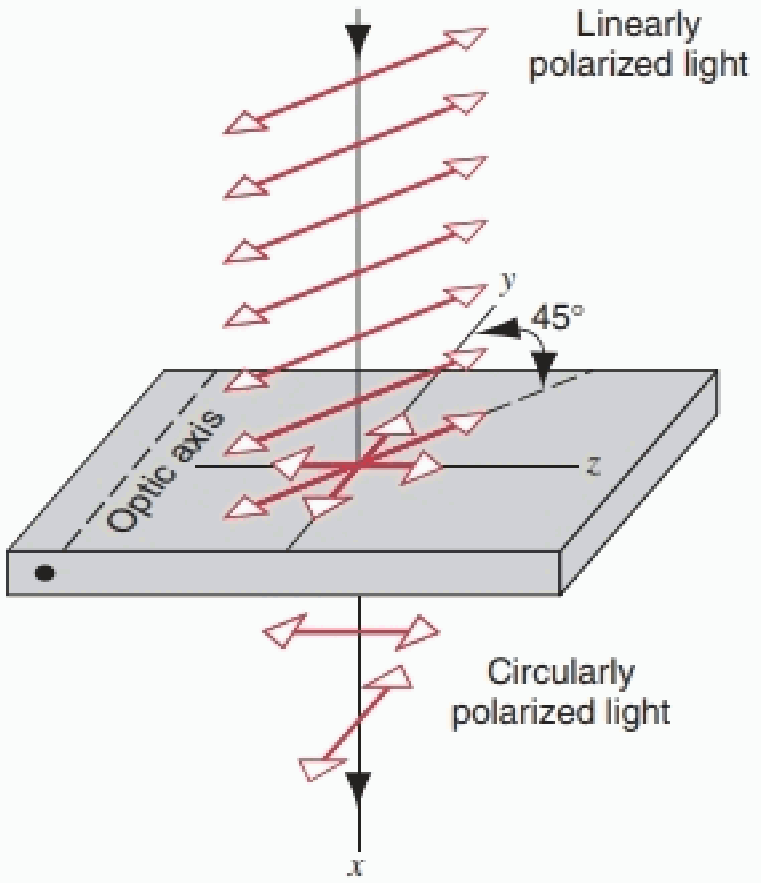


FIGURE 44-16. Linearly polarized light falls on a doubly refracting slab cut with its optic axis parallel to the surface. The plane of polarization makes an angle of 45° with the optic axis. Two polarization states that were in phase before entering the slab are out of phase when they emerge from this slab.

44-5 CIRCULAR POLARIZATION

Figure 44-16 shows a linearly polarized wave incident on a thin slab of a birefringent material such as calcite. The wave travels in the x direction, which is perpendicular to the face of the slab (the yz plane). The material has been cut so that the optic axis is parallel to the face of the slab. We choose the plane of polarization to be at an angle of 45° with the optic axis, so that the components of the $\vec{E}$ field along the axis and perpendicular to the axis have equal amplitudes.

As we illustrated in Fig. 44-15, the direction perpendicular to the optic axis is the one in which the speeds of the o-wave and e-wave differ by the greatest amount. We also know that the polarizations of the o-wave and e-wave are perpendicular to one another. If we resolve the incident wave into two waves, one polarized in the xy plane and one in the xz plane, these waves can represent the perpendicular polarization states of the o-wave and e-wave, and they travel at different speeds through the slab. The two polarization states shown below the slab in Fig. 44-16 represent $\vec{E}$ fields that were originally part of the same wavefront before passing through the slab; after passing through the slab, there is a phase difference between them that depends on the thickness of the slab and on its two indices of refraction.

Let us choose the thickness of the slab so that this phase difference is 90°; that is, one component has its maximum magnitude where the other compound has zero amplitude. (Because the indices of refraction vary with wavelength, this thickness will likewise vary with the wavelength of the light.) In this case the slab is called a *quarter-wave plate,* because 90° represents 1/4 of a full cycle of oscillation.

Figure 44-17 shows the two perpendicular polarization states 90° out of phase as they emerge from the quarter-wave plate. At any point, the net electric field is the vector sum of these two components. As the wave travels, the direction of this net electric field appears to rotate about the direction of propagation (the x axis in Fig. 44-17). From the perspective of a point on the x axis, the tip of this net $\vec{E}$ field traces out a circle about the x axis; for this reason this type of wave is called a *circularly polarized wave.* In contrast to a plane-polarized wave, in which the direction of polarization is fixed, here the direction of polarization changes in a continuous way as the wave travels. This rotation of the direction of polarization takes place with an angular frequency $\omega = 2\pi f$, where f is the frequency of the electromagnetic wave. Because light waves have frequencies in the range of 10^{15} Hz, this rotation is too rapid to be observed directly.

As the wave moves past an observer facing the oncoming light and located at a fixed point on the x axis in Fig. 44-17, the $\vec{E}$ vector appears to rotate clockwise, which is called *right-circular polarization.* By changing the orientation of the incident linear polarization relative to the optic axis in Fig. 44-16, we can produce a situation in which the $\vec{E}$ vector appears to rotate in the opposite direction, which is called *left-circular polarization.*

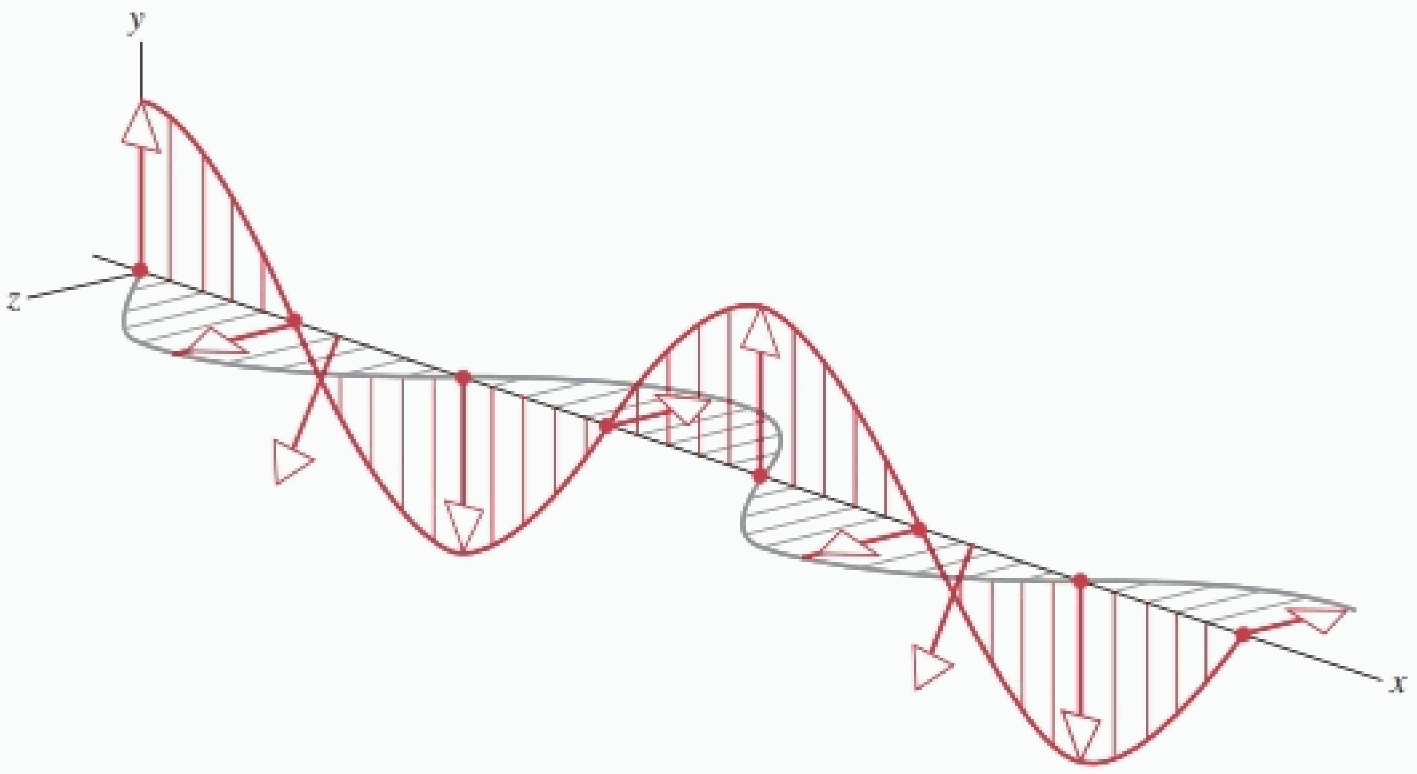


FIGURE 44-17. Two waves of equal amplitude and linearly polarized in perpendicular directions move in the x direction. Only the $\vec{E}$ vectors are shown. The waves differ in phase by 90°, such that one reaches its maximum when the other is zero. The resultant $\vec{E}$ vector appears to rotate clockwise with time to an observer at a fixed location on the x axis.

As in the case of unpolarized light, when circularly polarized light falls on a polarizing sheet, the emerging light is plane polarized and has half the intensity of the incident light. A polarizing sheet therefore cannot be used to distinguish between unpolarized and circularly polarized light. To make this distinction, we can simply reverse the experiment of Fig. 44-16: if circularly polarized light is incident on a quarter-wave plate, *another* 90° phase difference will be introduced, which (depending on the orientation of the axis of the quarter-wave plate) will either add to or subtract from the original 90° phase difference to give a phase difference of either 0 or 180°. As a result, the two components reach their maxima at the same time, and the emerging light is linearly polarized at an angle of ±45° with the optic axis. We could demonstrate this linear polarization by passing the light through a polarizing sheet and rotating the sheet to show that the intensity goes to zero when the axis of the sheet makes an angle of 45° with the optic axis of the quarter-wave plate.

SAMPLE PROBLEM 44-3. A quartz quarter-wave plate is to be used with sodium light ($\lambda = 589$ nm). What is the minimum thickness of such a plate?

Solution Two waves travel through the slab at speeds corresponding to the two principal indices of refraction given in Table 44-1 ($n_e = 1.553$ and $n_o = 1.544$). If the crystal thickness is x, the number of wavelengths of the first wave contained in the crystal is

$$N_e = \frac{x}{\lambda_e} = \frac{xn_e}{\lambda},$$

where λ_e is the wavelength of the e-wave in the crystal and λ is the wavelength in air. For the second wave, the number of wavelengths is

$$N_o = \frac{x}{\lambda_o} = \frac{xn_o}{\lambda},$$

where λ_o is the wavelength of the o-wave in the crystal. The difference $N_e - N_o$ must be $m + \frac{1}{4}$, where $m = 0, 1, 2, \ldots.$ The minimum thickness corresponds to $m = 0$, in which case

$$\frac{1}{4} = \frac{x}{\lambda}(n_e - n_o).$$

This equation yields

$$x = \frac{\lambda}{4(n_e - n_o)} = \frac{589 \text{ nm}}{(4)(1.553 - 1.544)} = 0.016 \text{ mm}.$$

This plate is rather thin. Most quarter-wave plates are made from mica; the sheet is split to the correct thickness by trial and error.

SAMPLE PROBLEM 44-4. A linearly polarized light wave of amplitude E_0 falls on a calcite quarter-wave plate with its plane of polarization at 45° to the optic axis of the plate, which is taken as the y axis; see Fig. 44-18. The emerging light will be circularly polarized. In what direction will the electric vector appear to rotate? The direction of propagation is out of the page.

Solution The wave component whose vibrations are parallel to the optic axis (the e wave) can be represented as it emerges from the plate as

$$E_y = (E_0 \cos 45°) \sin \omega t = \frac{1}{\sqrt{2}} E_0 \sin \omega t = E_m \sin \omega t.$$

The wave component whose vibrations are at right angles to the optic axis (the o-wave) can be represented as

$$E_z = (E_0 \sin 45°) \sin(\omega t - 90°) = -\frac{1}{\sqrt{2}} E_0 \cos \omega t$$
$$= -E_m \cos \omega t,$$

the 90° phase shift representing the action of the quarter-wave plate. Note that E_z reaches its maximum value one-fourth of a cycle *later* than E_y does, because, in calcite, wave E_z (the o-wave) travels *slower* than wave E_y (the e-wave).

To decide the direction of rotation, let us locate the tip of the rotating electric vector at two instants of time, (Fig. 44-18*a*) $t = 0$ and (Fig. 44-18*b*) a short time t_1 later chosen so that ωt_1 is a small angle. At $t = 0$ the coordinates of the tip of the rotating vector (see Fig. 44-18*a*) are

$$E_y = 0 \quad \text{and} \quad E_z = -E_m.$$

At $t = t_1$ these coordinates become, approximately,

$$E_y = E_m \sin \omega t_1 \approx E_m \omega t_1,$$
$$E_z = -E_m \cos \omega t_1 \approx -E_m.$$

Figure 44-18*b* shows that the vector representing the emerging circularly polarized light is rotating counterclockwise; by convention such light is called *left-circularly polarized,* because the observer is always considered to face the light source.

You should verify that if the plane of vibration of the incident light in Fig. 44-18 is rotated through ±90°, the emerging light will be *right-circularly polarized.*

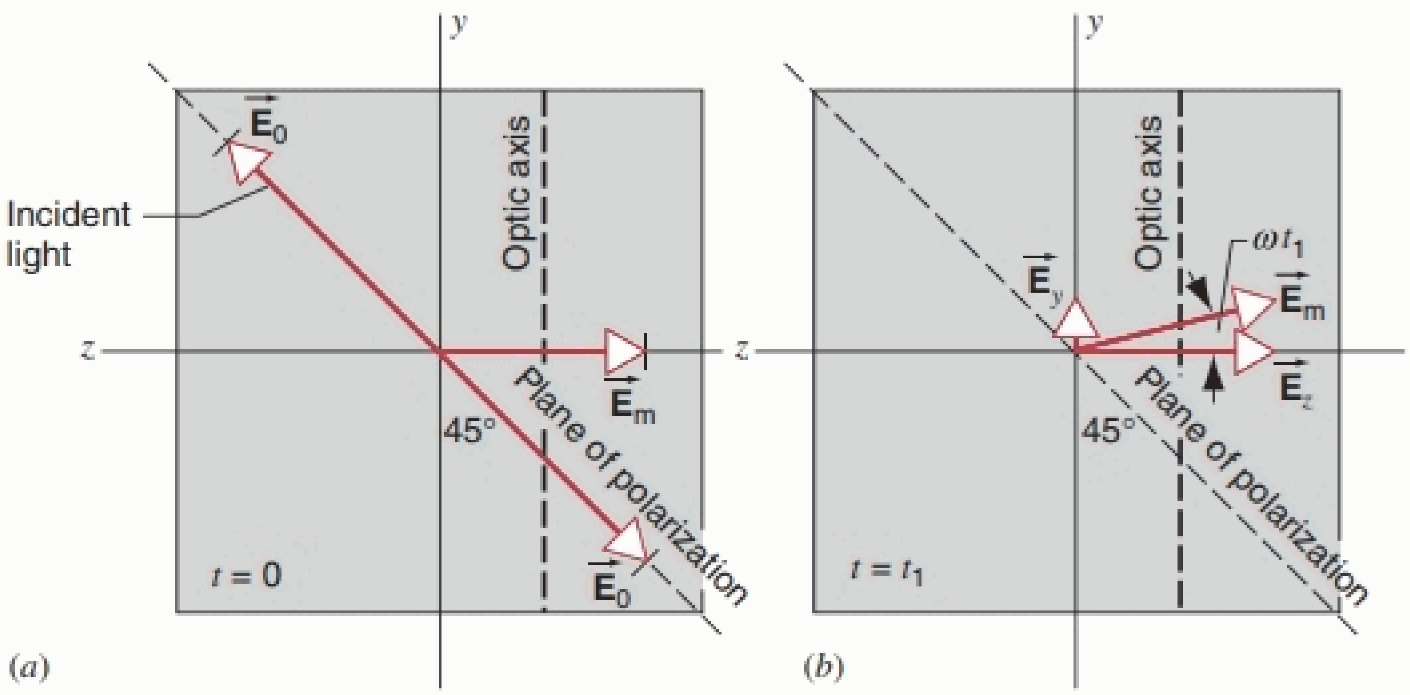


FIGURE 44-18. Sample Problem 44-4. Linearly polarized light falls (from behind the page) on a quarter-wave plate. The incident light is polarized at 45° with the y and z axes. (*a*) At a particular time $t = 0$, the emerging $\vec{\mathbf{E}}$ vector points in the $-z$ direction. (*b*) A short interval of time t_1 later the vector has rotated to a new position. In this case the $\vec{\mathbf{E}}$ vector rotates counterclockwise, as seen by an observer on the x axis facing the light source.

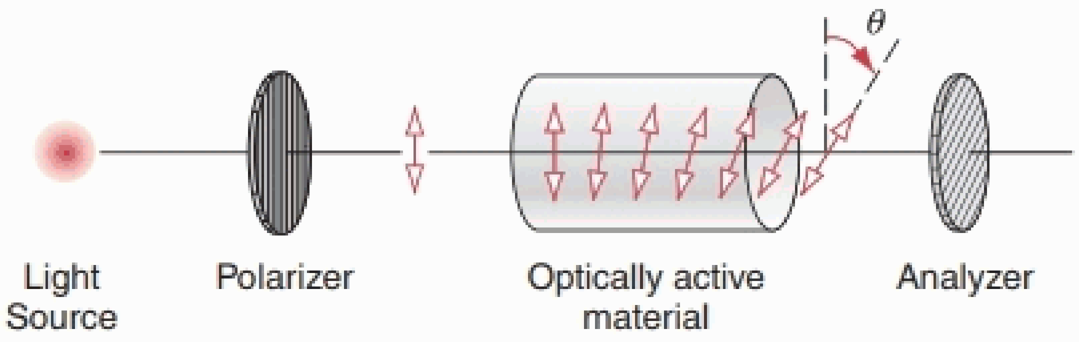


FIGURE 44-19. An optically active material rotates the direction of polarization of linearly polarized light through an angle θ.

Optical Activity

When linearly polarized light passes through certain materials, we find that the plane of polarization rotates about the direction of propagation through an angle that depends on the properties of the material and on the distance through the material that the light travels. This effect is called *optical activity,* and the materials are known as optically active. Quartz crystals and sugar solutions are examples of optically active materials. Figure 44-19 shows an example of how this rotation might be measured. The polarizer and analyzer can be carefully aligned when there is no material present between them; when the optically active material is present, the angle θ by which the analyzer must rotate for maximum transmission gives the rotation of the plane of polarization in passing through the material.

We can consider plane polarized light to be a superposition of left- and right-circularly polarized light. Imagine one $\vec{\mathbf{E}}$ vector rotating clockwise and another rotating counterclockwise, so that their vector sum (which is always along the bisector of the angle between them) simply oscillates back and forth along the same direction. In passing through an optically active material, one circular polarization component moves at a greater speed than the other. This causes a change in the relative phase of the two components, and as a result the direction of their vector sum is altered. This effect is directly proportional to the distance traveled through the material.

Many optically active materials are organic liquids. Organic compounds derived from plants and animals often have asymmetric helical structures, and when light passes through such materials the shape of the molecule enhances the passage of one circular polarization component relative to the other. On the other hand, it is possible to synthesize organic compounds from nonliving materials, in which case they contain equal amounts of the two asymmetric helical forms. Such materials are found not to be optically active.

44-6 POLARIZATION BY SCATTERING

A light wave, falling on a transparent solid, causes the electrons in the solid to oscillate periodically in response to the time-varying electric vector of the incident wave. The wave that travels through the medium is the resultant of the incident wave and the radiations from the oscillating electrons. The resultant wave has a maximum intensity in the direction of the incident beam, falling off rapidly on either side. The lack of sideways scattering, which would be essentially complete in a large "perfect" crystal, comes about because the oscillating charges in the medium act cooperatively or coherently.

When light passes through a liquid or a gas, we find much more sideways scattering. The oscillating electrons in this case, being separated by relatively large distances and not being bound together in a rigid structure, act independently rather than cooperatively. Thus a rigid cancellation of wave disturbances that are not in the forward direction is less likely to occur; there is more sideways scattering.

Light scattered sideways from a gas can be wholly or partially polarized, even though the incident light is unpolarized. Figure 44-20 shows an unpolarized beam moving upward on the page and striking a gas atom at O. The electrons at O oscillate in response to the electric components of the incident wave, their motion being equivalent to two oscillating dipoles whose axes are in the y and z directions at O. For transverse electromagnetic waves, an oscillating dipole does not radiate along its own axis. Thus an observer at O' would receive no radiation from the dipole at O oscillating in the z direction. The radiation reaching O' would come entirely from the dipole at O oscillating in the y direction and would be linearly polarized in the y direction.

As observer O' moves off the z axis, the radiation becomes less than fully polarized, because the dipole at O oscillating along the z axis can radiate somewhat in these directions. At points along the x axis, the transmitted ($x > 0$) or backscattered ($x < 0$) radiation is unpolarized, because both dipoles can radiate equally well in the x direction.

A familiar example of this effect is the scattering of sunlight by the molecules of the Earth's atmosphere. If the atmosphere were not present, the sky would appear black

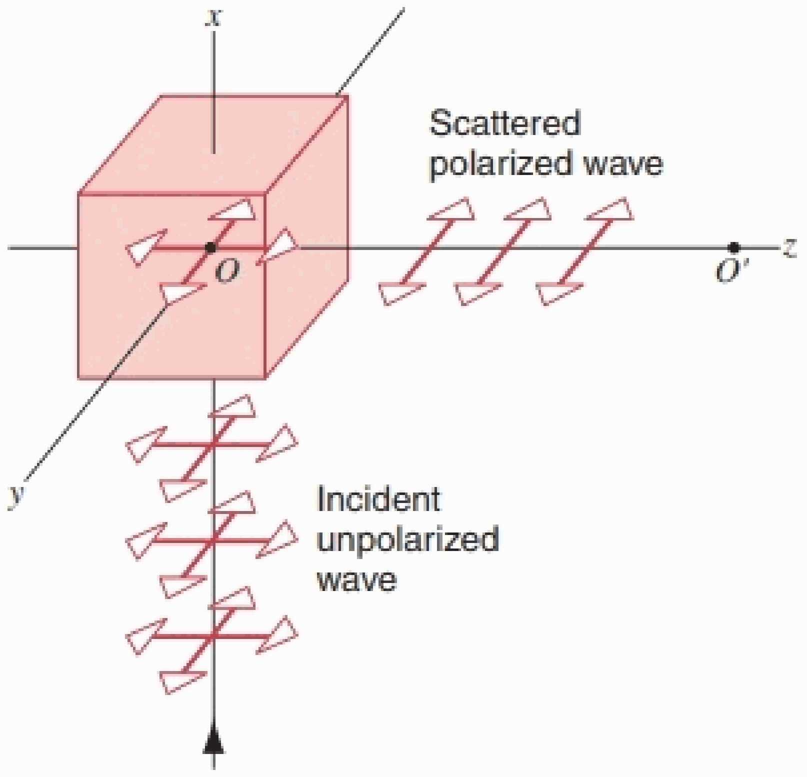


FIGURE 44-20. An unpolarized incident wave is scattered by an atom at O. The wave scattered toward O' on the z axis is linearly polarized.

except in the direction of the Sun, as observed by astronauts orbiting above the atmosphere. We can easily check with a polarizer that the light from the cloudless sky is at least partially polarized. This fact is used in polar exploration in the *solar compass.* In this device we establish direction by noting the nature of the polarization of the scattered sunlight. As is well known, magnetic compasses are not useful in these regions. It has been learned* that bees orient themselves in their flights between their hive and the pollen sources by means of polarization of the light from the sky; bees' eyes contain built-in polarization-sensing devices.

It still remains to be explained why the light scattered from the sky is predominantly blue and why the light received directly from the Sun—particularly at sunset when the length of the atmosphere that it must traverse is greatest—is red. The cross section of an atom or molecule for light scattering depends on the wavelength, blue light being scattered more effectively than red light. Since the blue light is more strongly scattered, the transmitted light has the color of normal sunlight with the blues largely removed; it is therefore more reddish in appearance.

The conclusion that the scattering cross section for blue light is higher than that for red light can be made reasonable with a mechanical analogy. An electron in an atom or molecule is bound there by strong restoring forces. It has a definite natural frequency, like a small mass suspended in space by an assembly of springs. The natural frequency for electrons in atoms and molecules is usually in a region corresponding to violet or ultraviolet light.

When light is allowed to fall on such bound electrons, it sets up forced oscillations at the frequency of the incident light beam. In mechanical resonant systems it is possible to "drive" the system most effectively if we apply an external force whose frequency is as close as possible to the natural resonant frequency. In the case of light, the frequency of blue light is closer to the natural resonant frequency of the bound electron than is that of red light. We would expect the blue light to be more effective in causing the electron to oscillate, and it is more effectively scattered.

Double Scattering (Optional)

Experiments similar to that shown in Fig. 44-20 can demonstrate that electromagnetic waves must be transverse; that is, there can be no component of the $\vec{\mathbf{E}}$ vector parallel to the direction of propagation. Suppose there were such a component along the direction of the incident wave (the x direction in Fig. 44-20). Then the electrons at O would oscillate in all three directions, and the scattered wave directed toward O' would show all three possible polarization directions (two transverse and one longitudinal). This radiation would thus be unpolarized. If the incident radiation is only transverse, as in Fig. 44-20, the radiation propagated to O' is linearly polarized. The question as to the transverse nature of the radiation is thus equivalent to determining whether the radiation traveling to O' is polarized or unpolarized.

There is another way to make this determination. Let us place a second scatterer at O'. A dipole at O' will oscillate in response to the incident (polarized) wave in only one direction (the y' direction, that of the incident $\vec{\mathbf{E}}$ vector, as shown in Fig. 44-21). Radiation scattered by that dipole can travel in the $\pm x'$ directions, but (for transverse radiation) *not in the* y' *direction.* Thus a detector D measuring the *intensity* of the radiation should see a maximum in the $\pm x'$ directions and a minimum of zero intensity in the $\pm y'$ direction. Such an experiment, as illustrated in Fig. 44-21, is called a *double scattering* experiment. Note that the *polarization* of the radiation scattered by the first target is determined through the *intensity* of the radiation scattered by the second target. If the radiation traveling to O' were not polarized (and not purely transverse), then the detector D would record the same intensity in all directions.

We can establish the transverse nature of electromagnetic radiation either by measuring the polarization of the radiation scattered from the first target (as shown in Fig. 44-20) or the intensity distribution of radiation scattered from the second target (as shown in Fig. 44-21). For some radiations (such as light), polarization measurements are relatively easy to make, and the double scattering method provides no great advantage. For other radiations (such as x rays or gamma rays), double scattering is usually the

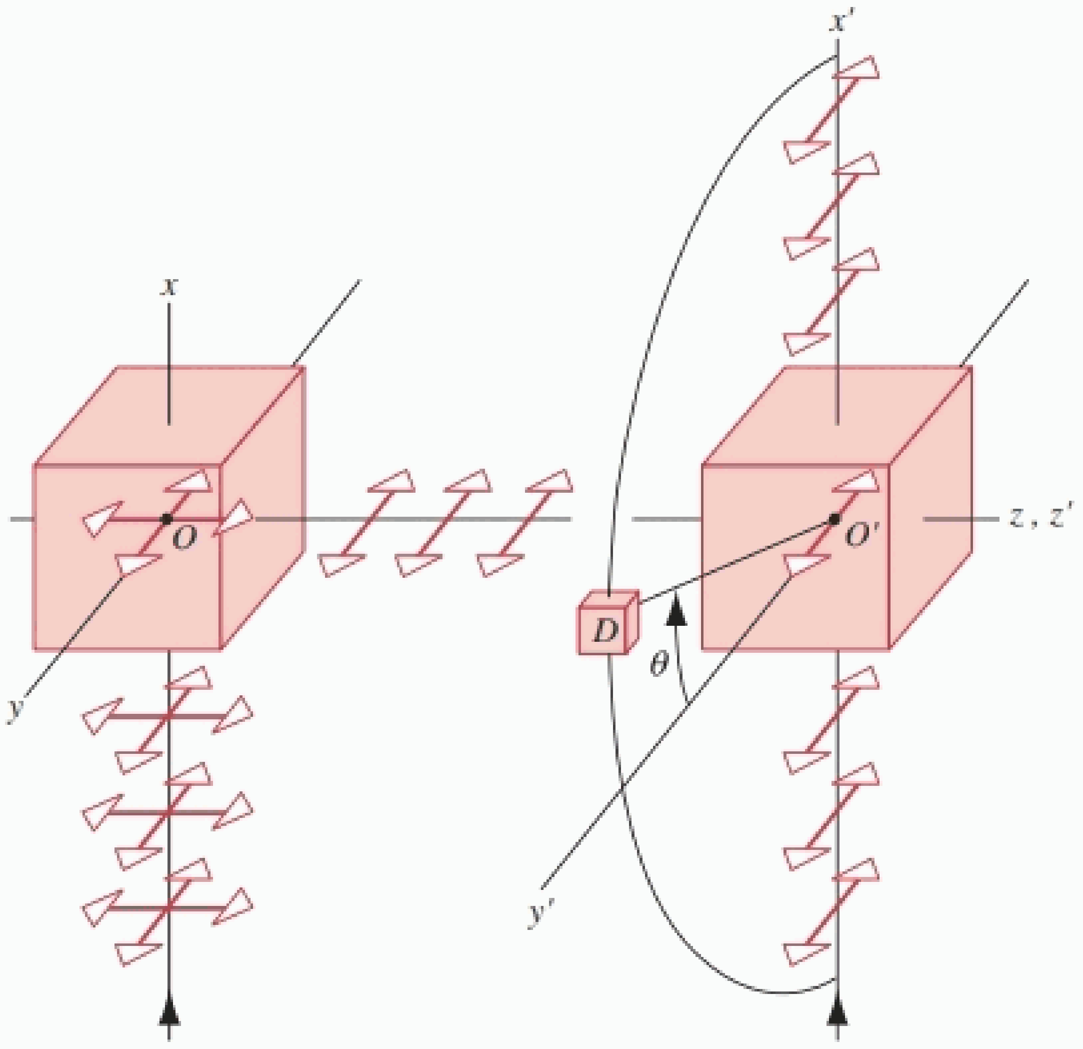


FIGURE 44-21. The polarized radiation scattered at O can be scattered by another atom at O'. A detector D measures the intensity of the radiation scattered by O' at various locations θ in the $x'y'$ plane.

---

* See "Polarized-Light Navigation by Insects," by Rudiger Wehner, *Scientific American,* July 1976, p. 106.

preferred method. Indeed, following the discovery of x rays in 1898, there was speculation on whether they were waves or particles. A double scattering experiment, performed in 1906 by Charles Barkla, established that x rays, like visible light, were transverse in nature and helped to confirm that x rays are part of the electromagnetic spectrum.

MULTIPLE CHOICE

44-1 Polarization of Electromagnetic Waves

44-2 Polarizing Sheets

1. A vertically oriented, ideal polarizing sheet transmits 50% of the incident unpolarized light. The polarizing sheet is now rotated through 45°. What fraction of the incident intensity now passes?

(A) 0% (B) 50% (C) 100%
(D) Either 0% or 100%

2. A vertically oriented, ideal polarizing sheet transmits 50% of the incident polarized light. The polarizing sheet is now rotated through 45°. What fraction of the incident intensity now passes?

(A) 0% (B) 50% (C) 100%
(D) Either 0% or 100%

3. An ideal polarizing sheet transmits 90% of the incident *partially* polarized light.

(*a*) What percentage of the incident light is polarized?

(A) Between 0% and 10%
(B) Between 0% and 20%
(C) Between 80% and 90%
(D) Between 80% and 100%
(E) 100%

(*b*) What fraction of the transmitted light is polarized?

(A) Between 0% and 10%
(B) Between 0% and 20%
(C) Between 80% and 90%
(D) Between 80% and 100%
(E) 100%

(*c*) The polarizing sheet is now rotated through 45°; during the rotation process the transmitted light monotonically decreases in intensity. What fraction of the incident intensity now passes?

(A) 0%
(B) Between 0% and 50%
(C) 50%
(D) Between 50% and 80%
(E) Between 40% and 90%

4. Two ideal polarizing sheets are stacked so that none of the incident *unpolarized* light is transmitted. A third polarizing sheet is slipped between the first two sheets at an angle of 45° to the bottom sheet. The fraction of light transmitted through the entire stack is

(A) still zero. (B) 1/8. (C) 1/4. (D) 1/2.

5. Ten polarizing filters are arranged so that the angle $\Delta\theta$ between the polarization direction of any two adjacent filters is the same. This angle is chosen so that each filter allows all but 10% of the light from the previous filter to pass through.

(*a*) The incident light is polarized with a polarization direction $\Delta\theta$ from the first sheet. What fraction of the light intensity passes through the entire stack?

(A) 0 (B) $(0.1)^{10}$ (C) $(0.9)^{10}$ (D) $(0.1)^9$

(*b*) The polarized incident light is replaced with unpolarized light. What fraction of the light intensity now passes through the entire stack?

(A) 0.05 (B) $\frac{1}{2}(0.1)^9$ (C) $\frac{1}{2}(0.9)^9$ (D) $(0.1)^9$

6. What would happen to the interference pattern produced from a twin-slit apparatus if the light from one slit is passed through a vertically polarizing sheet and the light through other slit is passed through a horizontally polarizing sheet?

(A) The interference pattern will look like a normal double-slit pattern.
(B) The interference pattern will be fainter than a normal double-slit pattern.
(C) The interference pattern will be more diffuse than a normal double-slit pattern.
(D) There will be no interference pattern.

44-3 Polarization by Reflection

7. Light reflected from the surface of a road is $\frac{1}{3}$ vertically polarized and $\frac{2}{3}$ horizontally polarized.

(*a*) At what angle should the polarization direction of a polarizing sheet be oriented to provide the maximum reduction in light intensity?

(A) vertically
(B) 30° from the vertical
(C) 30° from the horizontal
(D) horizontally

(*b*) What fraction of light passes through the filter if it is oriented to provide the maximum reduction in light intensity?

(A) $\frac{2}{3}$ (B) $\frac{1}{3}$ (C) $\frac{1}{6}$ (D) 0

8. Unpolarized light strikes an air–water interface at an angle so that the reflected ray is completely polarized as shown in Fig. 44-22*a*. A second ray of unpolarized light travels *backwards*, parallel to the previously refracted ray as shown in Fig. 44-22*b*. Describe the polarization of the reflected ray in Fig. 44-22*b*.

(A) The ray is completely polarized.
(B) The ray is partially polarized.
(C) The ray undergoes total internal reflection.
(D) There is not enough information to answer the question.

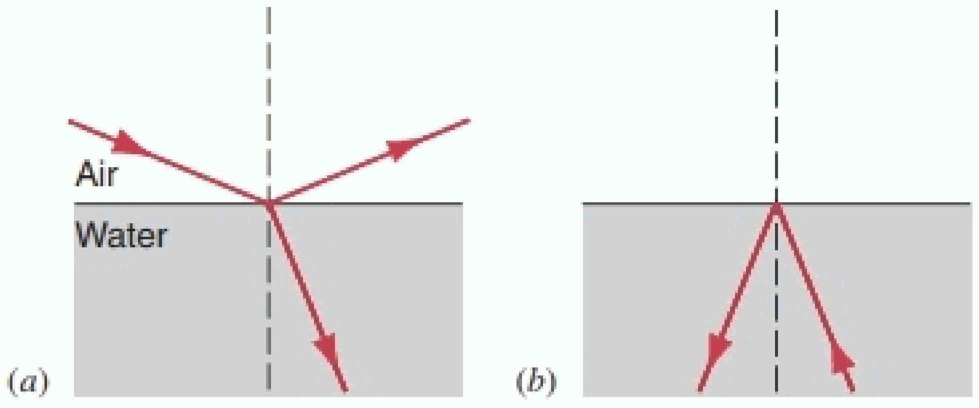


FIGURE 44-22. Multiple-choice question 8.

44-4 Double Refraction

44-5 Circular Polarization

9. Left-circularly polarized light is reflected from a plane mirror. What is the polarization state of the reflected ray?
   (A) Left-circularly polarized
   (B) Right-circularly polarized
   (C) Plane polarized
   (D) The answer depends on the angle of incidence.

44-6 Polarization by Scattering

10. Look straight up when the sun is on the horizon in the east. What polarization state is the blue light from the sky directly above you?
   (A) It is unpolarized.
   (B) It is primarily polarized in an east/west direction.
   (C) It is primarily polarized in a north/south direction.
   (D) There is not enough information to answer this question.

11. A beam of light passes through a polarizing filter before being projected horizontally into a dilute solution of milk in water. What orientation of the transmission axis of the polarizing sheet will maximize the intensity of the scattered light as viewed from above?
   (A) Vertical (B) Horizontal
   (C) The scattered intensity is independent of the polarizing direction.
   (D) There is not enough information to answer this question.

QUESTIONS

1. It is said that light from ordinary sources is unpolarized. Can you think of any common sources that emit polarized light?

2. Light from a laboratory gas discharge tube is unpolarized. How can this be made consistent with the fact that atoms and molecules radiate as electric dipoles whose radiation is linearly polarized?

3. Polarizing sheets contain long hydrocarbon chains that are made to line up in a parallel array during the production process. Explain how a polarizing sheet is able to polarize light. (Hint: Electrons are relatively free to move along these chains.)

4. As we normally experience them, radio waves are almost always polarized and visible light is almost always unpolarized. Why is this so?

5. What determines the desirable length and orientation of the rabbit ears on a portable TV set?

6. Why are sound waves unpolarized?

7. Why do sunglasses made of polarizing materials have a marked advantage over those that simply depend on absorption effects? What disadvantages might they have?

8. Unpolarized light falls on two polarizing sheets so oriented that no light is transmitted. If a third polarizing sheet is placed between them, can light be transmitted? If so, explain how.

9. Sample Problem 44-1 shows that, when the angle between the two polarizing directions is turned from 0° to 45°, the intensity of the transmitted beam drops to one-half its initial value. What happens to this "missing" energy?

10. You are given a number of polarizing sheets. Explain how you would use them to rotate the plane of polarization of a linearly polarized wave through any given angle. How could you do it with the least energy loss?

11. In the early 1950s, 3-D movies were very popular. Viewers wore polarizing glasses and a polarizing sheet was placed in front of each of the two projectors needed. Explain how the system worked. Can you suggest any problems that may have led to the early abandonment of the system?

12. A wire grid, consisting of an array of wires arranged parallel to one another, can polarize an incident unpolarized beam of electromagnetic waves that pass through it. Explain the facts that (*a*) the diameter of the wires and the spacing between them must be much less than the incident wavelength to obtain effective polarization and (*b*) the transmitted component is the one whose electric vector oscillates in a direction perpendicular to the wires.

13. Brewster's law, Eq. 44-2, determines the polarizing angle on reflection from a dielectric material such as glass; see Fig. 44-11. A plausible interpretation for zero reflection of the parallel component at that angle is that the charges in the dielectric are caused to oscillate parallel to the reflected ray by this component and produce no radiation in this direction. Explain this and comment on the plausibility.

14. Explain how polarization by reflection could occur if the light is incident on the interface from the side with the higher index of refraction (glass to air, for example).

15. Find a way to identify the polarizing direction of a polarizing sheet. No marks appear on the sheet.

16. Is the optic axis of a doubly refracting crystal simply a line or a direction in space? Has it a direction sense, like an arrow? What about the characteristic direction of a polarizing sheet?

17. If ice is doubly refracting (see Table 44-1), why do we not see two images of objects viewed through an ice cube?

18. Is it possible to produce interference effects between the *o*-beam and the *e*-beam, which are separated by the calcite crystal from the incident unpolarized beam in Fig. 44-14, by recombining them? Explain your answer.

19. From Table 44-1, would you expect a quarter-wave plate made from calcite to be thicker than one made from quartz?

20. Does the *e*-wave in doubly refracting crystals always travel at a speed given by c/n_e?

21. Devise a way to identify the direction of the optic axis in a quarter-wave plate.

22. If linearly polarized light falls on a quarter-wave plate with its plane of vibration making an angle of (*a*) 0° or (*b*) 90° with the axis of the plate, describe the transmitted light. (*c*) If this angle is arbitrarily chosen, the transmitted light is called elliptically polarized; describe such light.

23. You are given an object that may be (*a*) a disk of grey glass, (*b*) a polarizing sheet, (*c*) a quarter-wave plate, or (*d*) a half-wave plate (see Exercise 15). How could you identify it?

24. Can a linearly polarized light beam be represented as a sum of two circularly polarized light beams of opposite rotation? What effect does changing the phase of one of the circular components have on the resultant beam?

25. Could a radar beam be circularly polarized?

26. How can a right-circularly polarized light beam be transformed into a left-circularly polarized beam?

27. A beam of light is said to be unpolarized, linearly polarized, or circularly polarized. How could you choose among them experimentally?

28. A parallel beam of light is absorbed by an object placed in its path. Under what circumstances will (*a*) linear momentum and (*b*) angular momentum be transferred to the object?

29. When observing a clear sky through a polarizing sheet, you find that the intensity varies on rotating the sheet. This does not happen when viewing a cloud through the sheet. Why?

30. In 1949, it was discovered that light from distant stars in our galaxy is slightly linearly polarized, with the preferred plane of vibration being parallel to the plane of the galaxy. This is probably due to nonisotropic scattering of the starlight by elongated and slightly aligned interstellar grains (see Problem 16 in Chapter 22). If the grains are oriented with their long axes parallel to the interstellar magnetic field lines, as discussed in Section 44-2, and they absorb and radiate electromagnetic waves like the oscillating electrons in a radio antenna, how must the magnetic field be oriented with respect to the galactic plane?

31. Is polarization or interference a better test for identifying waves? Do they give the same information?

EXERCISES

44-1 Polarization of Electromagnetic Waves

1. The magnetic field equations for an electromagnetic wave in free space are $B_x = B \sin (ky + \omega t)$, $B_y = B_z = 0$. (*a*) What is the direction of propagation? (*b*) Write the electric field equations. (*c*) Is the wave polarized? If so, in what direction?

2. Prove that two linearly polarized light waves of equal amplitude, with their planes of vibration at right angles to each other, cannot produce interference effects. (Hint: Prove that the intensity of the resultant light wave, averaged over one or more cycles of oscillation, is the same no matter what phase difference exists between the two waves.)

44-2 Polarizing Sheets

3. A beam of unpolarized light of intensity 12.2 mW/m^2 falls at normal incidence on a polarizing sheet. (*a*) Find the maximum value of the electric field of the transmitted beam. (*b*) Calculate the radiation pressure exerted on the polarizing sheet.

4. Unpolarized light falls on two polarizing sheets placed one on top of the other. What must be the angle between the characteristic directions of the sheets if the intensity of the transmitted light is one-third the intensity of the incident beam? Assume that each polarizing sheet is ideal—that is, that it reduces the intensity of unpolarized light by exactly 50%.

5. Three polarizing plates are stacked. The first and third are crossed; the one between has its axis at 45° to the axes of the other two. What fraction of the intensity of an incident unpolarized beam is transmitted by the stack?

6. A beam of linearly polarized light strikes two polarizing sheets. The characteristic direction of the second is 90° with respect to the incident light. The characteristic direction of the first is at angle θ with respect to the incident light. Find angle θ for a transmitted beam intensity that is 0.100 times the incident beam intensity.

7. A beam of unpolarized light is incident on a stack of four polarizing sheets that are lined up so that the characteristic direction of each is rotated by 30° clockwise with respect to the preceding sheet. What fraction of the incident intensity is transmitted?

8. A beam of light is linearly polarized in the vertical direction. The beam falls at normal incidence on a polarizing sheet with its polarizing direction at 58.8° to the vertical. The transmitted beam falls, also at normal incidence, on a second polarizing sheet with its polarizing direction horizontal. The intensity of the original beam is 43.3 W/m^2. Find the intensity of the beam transmitted by the second sheet.

9. Suppose that in Exercise 8 the incident beam was unpolarized. What now is the intensity of the beam transmitted by the second sheet?

44-3 Polarization by Reflection

10. Light traveling in water of index of refraction 1.33 is incident on a plate of glass of index of refraction 1.53. At what angle of incidence is the reflected light completely linearly polarized?

11. (*a*) At what angle of incidence will the light reflected from water be completely polarized? (*b*) Does this angle depend on the wavelength of the light?

12. When red light in vacuum is incident at the polarizing angle on a certain glass slab, the angle of refraction is 31.8°. What are (*a*) the index of refraction of the glass and (*b*) the polarizing angle?

13. Calculate the range of polarizing angles for white light incident on fused quartz. Assume that the wavelength limits are 400 and 700 nm and use the dispersion curve of Fig. 39-11.

44-4 Double Refraction

14. Linearly polarized light of wavelength 525 nm strikes, at normal incidence, a wurzite crystal that is cut with its faces parallel to the optic axis. What is the smallest possible thickness of the crystal if the emergent *o*- and *e*-rays combine to form linearly polarized light? See Table 44-1.

44-5 Circular Polarization

15. What would be the action of a half-wave plate (that is, a plate twice as thick as a quarter-wave plate) on (*a*) linearly polarized light (assume the plane of vibration to be at 45° to the optic axis of the plate), (*b*) circularly polarized light, and (*c*) unpolarized light?

16. Find the greatest number of quarter-wave plates, to be used with light of wavelength 488 nm, that could be cut from a dolomite crystal 0.250 mm thick.

44-6 Polarization by Scattering

PROBLEMS

1. At a particular beach on a particular day near sundown the horizontal component of the electric field vector is 2.3 times the vertical component. A standing sunbather puts on polarizing sunglasses; the glasses suppress the horizontal field component. (*a*) What fraction of the light energy received before the glasses were put on now reaches the eyes? (*b*) The sunbather, still wearing the glasses, lies on his side. What fraction of the light energy received before the glasses were put on reaches the eyes now?

2. A beam of light is a mixture of polarized and unpolarized light. When it is sent through a polarizing sheet, we find that the transmitted intensity can be varied by a factor of five depending on the orientation of the polarizing sheet. Find the relative intensities of these two components of the incident beam.

3. A stack of polarizing sheets is arranged so that the angle between any two adjacent sheets is α. The sheets are arranged so that N sheets rotate the plane of polarization by θ, where $\theta = N\alpha$. Calculate the fraction of light that will pass through the stack in the limit as $N \to \infty$. Assume that θ is fixed, so $\alpha \to 0$.

4. It is desired to rotate the plane of vibration of a beam of polarized light by 90°. (*a*) How might this be done using only polarizing sheets? (*b*) How many sheets are required for the total intensity loss to be less than 5.0%?

5. A polarizing sheet and a quarter-wave plate are glued together in such a way that, if the combination is placed with face *A* against a shiny coin, the face of the coin can be seen when illuminated by light of appropriate wavelength. When the combination is placed with face *A* away from the coin, the coin cannot be seen. (*a*) Which component is on face *A* and (*b*) what is the relative orientation of the components?

6. A narrow beam of unpolarized light falls on a calcite crystal cut with its optic axis as shown in Fig. 44-23. (*a*) For $t = 1.12$ cm and for $\theta_i = 38.8°$, calculate the perpendicular distance between the two emerging rays x and y. (*b*) Which is the *o*-ray and which is the *e*-ray? (*c*) What are the states of polarization of the emerging rays? (*d*) Describe what happens if a polarizer is placed in the incident beam and rotated. (Hint: Inside the crystal the $\vec{\mathbf{E}}$-vector vibrations for one ray are always perpendicular to the optic axis and for the other ray they are always parallel. The two rays are described by the indices n_o and n_e; *in this plane,* each ray obeys Snell's law.)

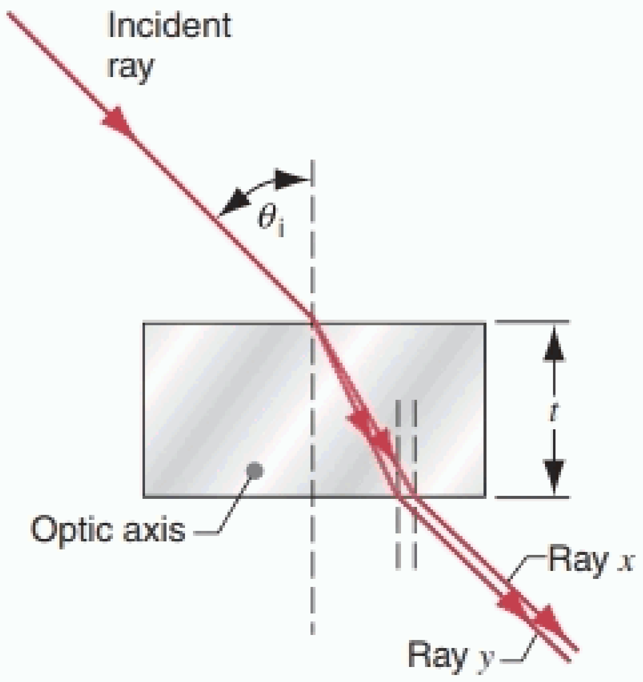


FIGURE 44-23. Problem 6.

7. A prism is cut from calcite so that the optic axis is parallel to the prism edge as shown in Fig. 44-24. Describe how such a prism might be used to measure the two principal indices of refraction for calcite. (Hint: See hint in Problem 6; see also Sample Problem 39-4.)

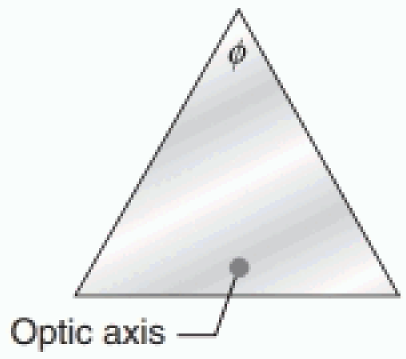


FIGURE 44-24. Problem 7.

CHAPTER 45

THE NATURE OF LIGHT

In this chapter, the first of eight devoted to quantum physics, we focus our attention on the nature of light. In Chapter 41 we saw the compelling evidence that light can be viewed as an electromagnetic wave. Nothing in this chapter detracts from that point of view or invalidates its supporting experiments. However, we introduce a second point of view—equally well buttressed by experiment—that light can also be viewed as a particle, called a photon.

We point out that both wave and particle are purely classical concepts and, as such, they cannot be extended deeply into the quantum realm. The wave picture holds for one large set of experiments, and the particle picture holds for another equally large and equally valid set. To learn the true nature of light we must dig deeper.

45-1 INTRODUCING THE PHOTON

In Section 41-2—which you should reread—we saw that Thomas Young's 1801 double-slit experiment provides strong evidence that visible light is wave-like. In this important experiment, light from a single source falls on a barrier containing two slits. The light flaring out from each of these slits then falls on a detector screen, which might be a photographic film. At certain points on the film the overlapping light beams from the two slits tend to reinforce each other and at other points they tend to cancel, forming a set of interference fringes like those shown in Fig. 41-4. This phenomenon of *interference* is a sure indicator that we are dealing with a wave.

We use the word "light" to represent not only visible light but radiations in all parts of the electromagnetic spectrum, from long radio waves to x rays and gamma rays. We can demonstrate by various kinds of interference experiments that all the radiations in this spectrum have a wave-like aspect.

In Section 43-4, for example, we described an interference experiment in which we allowed a beam of x rays to fall on a crystal. The atoms that make up the crystal form a periodic three-dimensional array of scattering centers for the incident x-ray beam. In certain directions—and in those directions only—the x rays scattered from the individual atoms interfere constructively to form intense, sharply defined rays. These rays can produce a characteristic pattern on a photographic film placed to intercept them; see Fig. 43-14. This experiment is strong evidence that x rays, like visible light, are wave-like.

There is, however, another property of light that seems at odds with the "light is a wave" concept. When light falls on a photographic film, energy is absorbed by individual silver halide grains of the emulsion, one grain at a time. It is not possible to predict at any instant which grain will be the next to absorb its fixed packet of energy. If light falls on a photoelectric device such as a photodiode, the output of the device can be viewed on an oscilloscope screen as a series of discrete pulses, randomly spaced in time. Alternatively, if the photodiode is connected to a speaker, its output can be heard as a series of randomly spaced "clicks."

When light falls on a detector such as a photographic film or a photoelectric device, the incident light does not transfer energy to the detector uniformly over the surface on which it falls but only at discrete, randomly located points. Furthermore, if the incident light is monochromatic—that is, if it consists of a single wavelength only—energy is transferred from the light beam to the detector not only at discrete points but also in a sharply defined discrete amount at each point. This discrete behavior suggests a particle-like, rather than a wave-like, nature for the incident light.

Light as Photons

In 1905, Einstein, building on the pioneering work of Max Planck* proposed that light could indeed exist as discrete bundles of energy, which we now call *photons.* According to Einstein, the energy E associated with a single photon is

$$E = hf. \tag{45-1}$$

In 1917, Einstein went further and assigned to each photon a linear momentum of magnitude

$$p = h/\lambda. \tag{45-2}$$

In these equations f and λ are, respectively, the frequency and the wavelength of the incident light, and h is a constant whose value is, to four significant figures,

$$h = 6.626 \times 10^{-34}\ \text{J}\cdot\text{s}.$$

The *Planck constant,* as this quantity is now called, proves to be the fundamental constant of quantum physics.

A glance at Eqs. 45-1 and 45-2 suggests the subtlety of the nature of light. On the left sides of these equations we have energy and momentum, properties that we associate with particles. On the right sides we have frequency and wavelength, properties that we associate with waves.

Both "particle" and "wave" are concepts that the human mind has developed over many millennia from familiarity with objects such as tossed pebbles on the one hand and water waves on the other. There is no way to extend such classical concepts unchanged to the quantum world. Do not try! In the quantum realm, physicists have learned to avoid all such concrete mental pictures unless the nature of an experiment makes it helpful to do so.

The Double-Slit Experiment Revisited

Let us return to the double-slit experiment of Section 41-2. We will detect the interference pattern by replacing the screen with a photographic film, and we will make the light source so weak that the total light energy in the apparatus at any time is no greater than the energy of a single photon. If we could examine the film minute by minute, we would see a pattern of dots that at first seems to be randomly distributed, but as more and more dots appear, the pattern of constructive and destructive interference fringes gradually takes shape.

An individual grain of the film is darkened by an interaction with a single photon. If we regard our light source as so weak that it emits individual photons one at a time, then a photon leaves the light source and later arrives at the film. Based on our experience with classical physics, we might then be tempted to ask how a photon gets from the source to the film. What is its trajectory? More specifically, through which slit did it pass?

If we try to answer such a question, we are clearly talking about the particle nature of light. Particles have trajectories; waves do not. If we regard light as individual concentrated bundles of energy, clearly they must each pass through one slit or the other. However, as we discussed in Chapter 41, the double-slit experiment depends on light behaving as a wave, so that we can divide the wavefront into two pieces and then recombine the two pieces. These two descriptions, one based on particles and the other on waves, are clearly inconsistent with one another. Is light a particle or a wave?

This puzzling question has two equally puzzling answers: (*a*) light is *neither* a particle nor a wave, and (*b*) light is *both* a particle and a wave. By the first answer, we mean that light does not satisfy the exclusive classical notions of strictly particle-like or strictly wave-like behavior. By the second answer, we mean that the particle and wave natures are both contained in what we call light, which can reveal one type of behavior or the other depending on the kind of experiment we do. Light has a more mysterious and complex behavior than is suggested by our simple ideas of "wave" or "particle."

In the next chapter we will find that electrons and other particles also have this complex dual nature. In this chapter we first discuss the experiments that suggest the photon nature of light, and then we consider a single apparatus that, at the whim of the experimenter, can reveal either the particle or wave nature of light that is already present in the apparatus.

45-2 THERMAL RADIATION

The concept of the photon was introduced into physics by studies of the radiation emitted by heated objects. These studies date from the middle of the 19th century and culminate in 1905, when Einstein introduced the idea of the photon. In this section we will review this development, which occurred during a period when the concepts of thermodynamics were being established. Indeed, these two lines of development were closely intertwined.

We see most objects by light that is reflected from them. However, objects also *emit* electromagnetic radiation and, if their temperatures are high enough, we can see, or otherwise detect, this emitted radiation. For example, the warmth that we feel when we put our hand near a lightbulb or a fire is due to the infrared radiation that is being emitted. Radiation emitted by a body because of its temperature is called *thermal radiation.* Such radiation is distributed continuously in wavelength, having measurable intensity throughout a broad range of the electromagnetic spectrum.

---

*Max Planck (1858–1947) was a German theoretical physicist whose specialization in thermodynamics led him to the study of thermal radiation and the discovery of the quantization of energy, for which he was awarded the 1918 Nobel Prize in physics.

The 19th-century investigators discovered two important properties of thermal radiation: (1) *As the temperature of a body is increased, the intensity of the emitted thermal radiation rises rapidly.* (2) *The higher the temperature, the shorter the wavelength of the most intense part of the radiant spectrum.* Attempts to express these observations in quantitative terms encountered a difficulty—namely, that the nature of the thermal radiation from a given body at a given temperature depends on the material of which the body is made and on such factors as its surface roughness. No fundamental law of physics seemed to be lurking behind such a large quantity of material-sensitive experimental data.

The solution to this difficulty was proposed in 1859 when Gustav Kirchhoff (German, 1824–1887) suggested that if you formed a cavity in a solid body, held its walls at a uniform temperature T, and drilled a small hole in its wall, the radiation emerging from the hole should not depend on the material or the mode of construction of the cavity but only on the temperature. Figure 45-1 shows the distribution in wavelength of such *cavity radiation,* as it is called, for two cavity temperatures. The cavity radiation (radiation in a box), because of its relative simplicity, plays the same role in thermal studies that the ideal gas (matter in a box) plays in the study of gases.

We can now express the first of the above two qualitative statements quantitatively. We can write the radiant intensity $I(T)$ (total power radiated at all wavelengths per unit surface area) of the cavity radiation as

$$I(T) = \sigma T^4, \tag{45-3}$$

in which T is the absolute (kelvin) temperature and σ is a universal constant called the *Stefan–Boltzmann constant*; its presently accepted value is $5.670 \times 10^{-8}\ \mathrm{W/m^2 \cdot K^4}$. Equation 45-3 is known as the *Stefan–Boltzmann law.* Josef Stefan (Austrian, 1835–1893) deduced this law in 1879 from a series of measurements; Boltzmann derived it theoretically a few years later, using concepts from thermodynamics and electromagnetic theory. The intensity $I(T)$ represents the *total* radiated power, summed or integrated over all wavelengths. In Fig. 45-1, the intensity for a given temperature is represented by the total area under the corresponding data points; these areas clearly increase markedly with temperature.

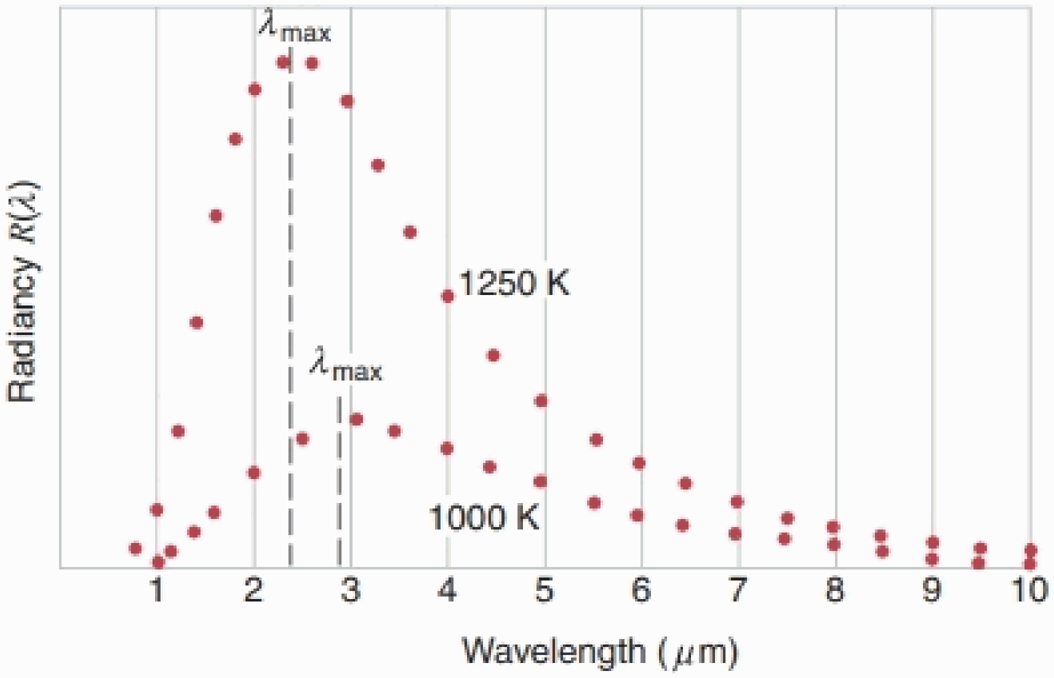


FIGURE 45-1. A possible result of the measurement of the radiancy over many different wavelengths for two different temperatures. The peak $\lambda_{\max}$ shifts to shorter wavelengths as the temperature increases.

We can express the second of our above qualitative statements by

$$\lambda_{\max} T = \text{a constant}, \tag{45-4}$$

in which $\lambda_{\max}$ is the wavelength at which the radiation emitted by a cavity at temperature T has its maximum intensity. The constant in Eq. 45-4 has the presently accepted value of $2898\ \mu\mathrm{m \cdot K}$. Equation 45-4 was deduced by Wilhelm Wien (German, 1864–1928) from thermodynamic principles and is called *Wien's displacement law.* This law also is evident in Fig. 45-1 in that the higher the temperature, the shorter is the wavelength at which the radiation is most intense. This law is consistent with familiar experience. As we heat a piece of metal in a fire, it first begins to glow a dull red and then, as its temperature increases, its color shifts to smaller wavelengths and it glows yellow-orange. Wien's law can be used to deduce the surface temperature of stars from their color, blue stars being hotter than red stars, for example.

The Stefan–Boltzmann law and Wien's law do not tell the whole story about thermal radiation. What is missing is an expression for the intensity of the cavity radiation as a function of wavelength—that is, an equation that fits the individual data points of Fig. 45-1. We seek to find an expression for the *spectral radiancy* $R(\lambda, T)$, which is defined such that the product $R(\lambda, T)\, d\lambda$ gives the emitted power per unit area whose wavelengths lie in the range λ to $\lambda + d\lambda$. The intensity $I(T)$ is related to the spectral radiancy by integrating R over the entire range of wavelengths; that is,

$$I(T) = \int_0^\infty R(\lambda, T)\, d\lambda, \tag{45-5}$$

in which the integration is carried over all wavelengths. Equation 45-5 shows that $I(T)$ is represented graphically in Fig. 45-1 by the area under the curve that connects the points at a particular temperature T.

During the late 1800s, a number of suggestions for the function $R(\lambda, T)$ were put forward. None agreed with the experimental data at all wavelengths. In 1900, Planck proposed an expression, written in modern notation as

$$R(\lambda, T) = \frac{2\pi c^2 h}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1} \tag{45-6}$$

and now known as *Planck's radiation law.* In this equation, c is the speed of light, k is the Boltzmann constant, and h is

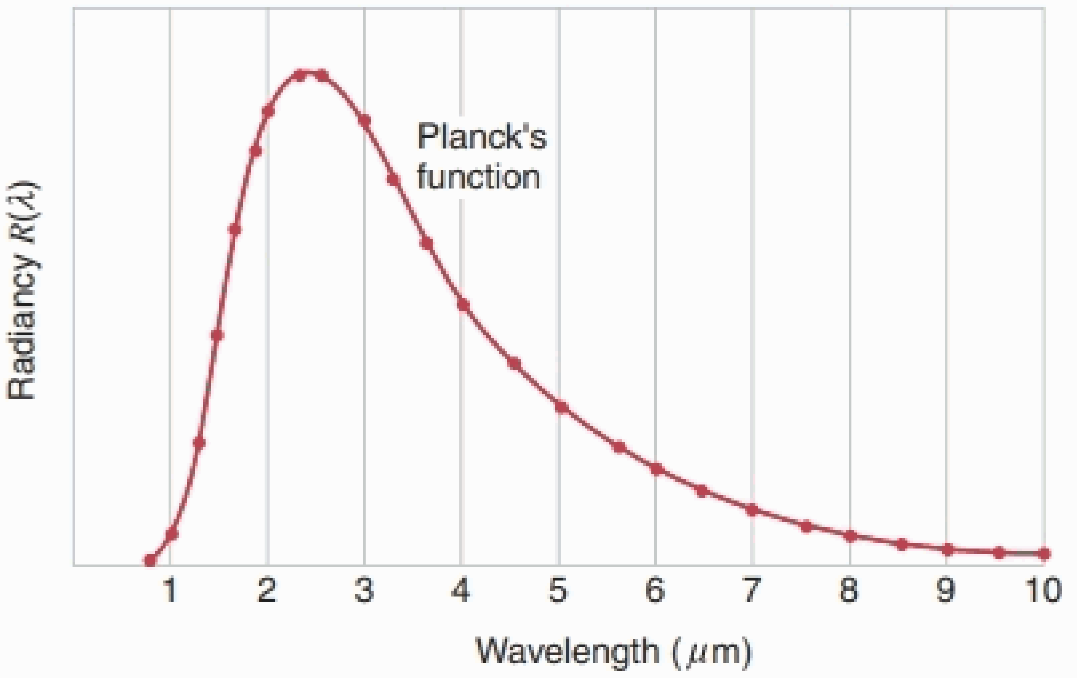


FIGURE 45-2. Planck's function fits the observed data perfectly.

a new constant, introduced into physics for the first time and now familiar to us as the Planck constant. Equation 45-6 fits the experimental data at all wavelengths and at all temperatures at which it has been tested, as indicated in Fig. 45-2. In 1990, for example, measurements of the cosmic background radiation, presumably remnants of the big bang, were obtained by the Cosmic Background Explorer (COBE) satellite and found to follow Planck's law with astonishingly good precision, the temperature deduced from the agreement being 2.725 ± 0.002 K (as we discuss in Chapter 52).

By substituting Eq. 45-6 into Eq. 45-5 and carrying out the integration, one can derive Eq. 45-3, the Stefan–Boltzmann law, and can show that the constant σ in that law is related to other fundamental constants by

$$\sigma = \frac{2\pi^5 k^4}{15c^2h^3}. \tag{45-7}$$

Using the best experimental value of the Stefan–Boltzmann constant available in 1900, Planck was able to obtain a value for his new constant of $h = 6.56 \times 10^{-34}\ \mathrm{J\cdot s}$, which differs by only about 1% from the current value. Furthermore, by differentiating Eq. 45-6, seeking the value of λ for which the spectral radiancy has its maximum value, one can derive Eq. 45-4, the Wien displacement law, and can show that the constant in that equation is related to other fundamental constants by

$$\lambda_{\max}T = \frac{hc}{4.9651k}, \tag{45-8}$$

which agrees with Eq. 45-4 and which reproduces the experimental value of the constant in that equation. For their pioneering discoveries in thermal radiation that laid the groundwork for the quantum theory, Planck (1918) and Wien (1911) were awarded the Nobel Prize in physics.

In deriving Eq. 45-6, Planck assumed that the atoms that formed the walls of the cavity behaved like tiny harmonic oscillators. He further assumed that these oscillators could not oscillate with any arbitrary energy, but only with energies that were integral multiples of hf, where f is the frequency of the radiation that these oscillators absorb and emit as they interact with the radiation trapped in the cavity. This astonishing assumption introduced quantization and the Planck constant into physics and was the beginning of the full development of quantum physics as we know it today. Physicists of that day, including Planck himself, found it difficult to believe that an oscillator could not absorb or emit energy in *any* amount within a reasonable range. In 1905, Einstein, bringing his great insight to bear on Planck's derivation, rederived Eq. 45-6 with a different basic assumption: that the radiation in the cavity, in its interaction with the walls of the cavity, behaved as if it consisted of energy quanta in amount given by Eq. 45-1—namely, $E = hf$. Thus Einstein introduced *the photon,* as these light quanta were later called, into physics. In 1916, Einstein gave a further derivation of Eq. 45-6; its simplicity and straightforward character further reinforced the photon concept.

In the next few sections we will consider the experimental evidence that supports the reality of Einstein's photons.

SAMPLE PROBLEM 45-1. How hot is a star? The "surfaces" of stars are not sharp boundaries like the surface of the Earth. Most of the radiation that a star emits is in thermal equilibrium with the hot gases that make up the star's outer layers. Without too much error, then, we can treat starlight as cavity radiation. Here are the wavelengths at which the spectral radiancies of three stars have their maximum values:

Star	$\lambda_{\max}$	*Appearance*
Sirius	240 nm	Blue-white
Sun	500 nm	Yellow
Betelgeuse	850 nm	Red

(*a*) What are the surface temperatures of these stars? (*b*) What are the radiant intensities of these three stars? (*c*) The radius r of the Sun is 7.0×10^8 m and that of Betelgeuse is over 500 times larger, or 4.0×10^{11} m. What is the total radiated power output (that is, the *luminosity* L) of these stars?

Solution (*a*) From Eq. 45-4 we find, for Sirius, (with $\lambda_{\max} = 240$ nm $= 0.240\ \mu$m)

$$T = \frac{2898\ \mu\mathrm{m\cdot K}}{\lambda_{\max}} = \frac{2898\ \mu\mathrm{m\cdot K}}{0.240\ \mu\mathrm{m}} = 12{,}000\ \mathrm{K}.$$

The temperatures for the Sun and for Betelgeuse work out in the same way to be 5800 K and 3400 K, respectively. At 5800 K, most of the radiation from the Sun's surface lies within the visible region of the spectrum. This suggests that over ages of evolution, eyes have adapted to the Sun to become most sensitive to those wavelengths that it radiates most intensely.

(*b*) For Sirius we have, from the Stefan–Boltzmann law (Eq. 45-3),

$$\begin{aligned} I = \sigma T^4 &= (5.67 \times 10^{-8}\ \mathrm{W/m^2\cdot K^4})(12{,}000\ \mathrm{K})^4 \\ &= 1.2 \times 10^9\ \mathrm{W/m^2}. \end{aligned}$$

The radiant intensities for the Sun and for Betelgeuse work out to be $6.4 \times 10^7\ \mathrm{W/m^2}$ and $7.7 \times 10^6\ \mathrm{W/m^2}$, respectively.

(*c*) We find the luminosity of a star by multiplying its radiant intensity by its surface area. Thus, for the Sun,

$$\begin{aligned} L &= I(4\pi r^2) = (6.4 \times 10^7\ \text{W/m}^2)(4\pi)(7.0 \times 10^8\ \text{m})^2 \\ &= 3.9 \times 10^{26}\ \text{W}. \end{aligned}$$

For Betelgeuse, the luminosity works out to be 1.5×10^{31} W, about 38,000 times larger. The enormous size of Betelgeuse, which is classified as a "red giant," much more than makes up for the relatively low radiant intensity associated with its low surface temperature.

The colors of stars are not strikingly apparent to the average observer because the retinal cones, which are responsible for color vision, do not function well in dim light. If this were not so, the night sky would be spangled with color.

45-3 THE PHOTOELECTRIC EFFECT

If we allow monochromatic light to fall on a metallic surface, electrons may be ejected from that surface. Figure 45-3 shows the experimental arrangement for studying this *photoelectric effect.* If we set up a suitable potential difference ΔV between the surface E (the emitter) and surface C (the collector), we can sweep up these ejected electrons and measure them as a photoelectric current i in the external circuit.

Figure 45-4 (curve a) shows the photoelectric current as a function of the potential difference ΔV. We see that, if the collector is at a higher potential than the emitter and if ΔV is large enough, the photoelectric current reaches a constant saturation value at which *all* electrons emerging from the emitter are collected.

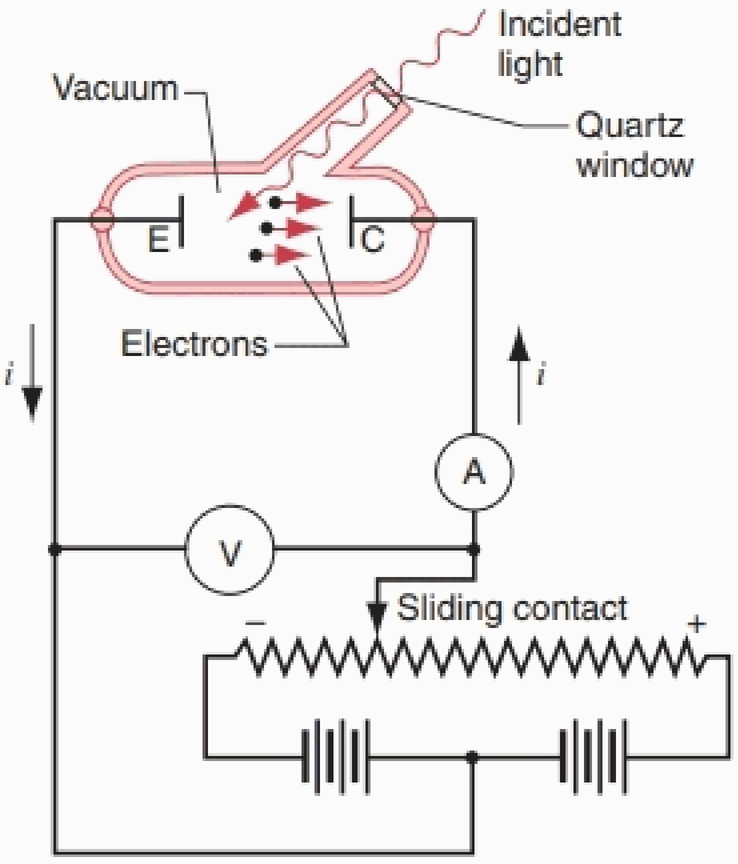


FIGURE 45-3. An apparatus for studying the photoelectric effect. The arrows show the direction of the photoelectric current in the external circuit, which is opposite to the motion of the (negatively charged) electrons. The effective potential difference between the emitter E and the collector C, which is read by the voltmeter V, is the algebraic sum of the potential difference associated with the batteries and that associated with the contact potential difference between the emitter and the collector. Because these components are made of different materials they form a "battery" in their own right.

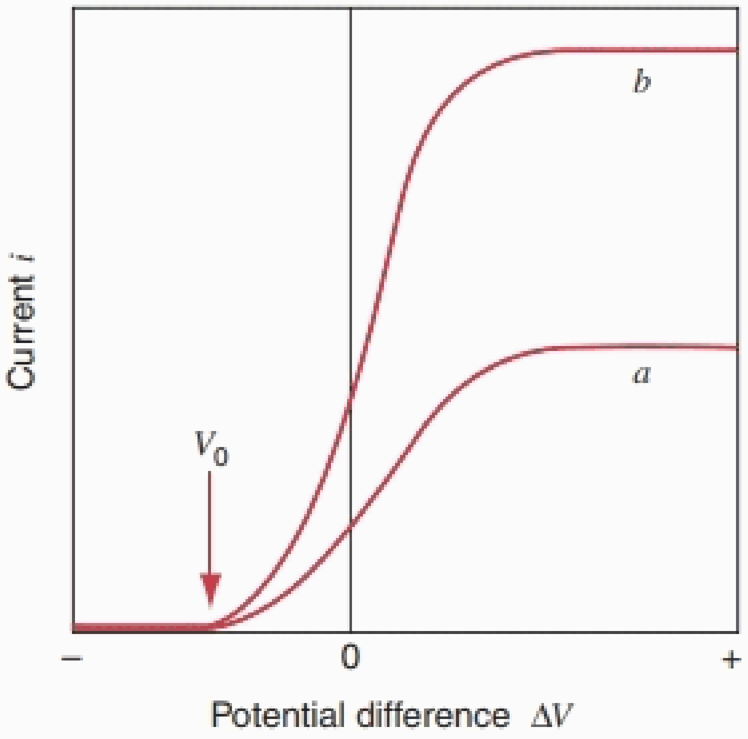


FIGURE 45-4. A plot (not to scale) of data taken with the apparatus of Fig. 45-3. The intensity of the incident light is twice as great for curve b as for curve a. The emitter and the wavelength of the incident light are the same for both runs. Note that the stopping potential is the same for each run.

If we reduce ΔV to zero, the photoelectric current does not drop to zero because the electrons are emitted with a definite range of speeds. However, if we reverse the sign of the potential difference and make ΔV large enough, we eventually reach a point at which even the most energetic emitted electrons are turned back before they strike the collector, and the photoelectric current i does indeed drop to zero. The magnitude of this (negative) potential difference is called the *stopping potential* V_0; that is, $V_0 = |\Delta V|$ at the point at which the photoelectric current drops to zero (Fig. 45-4).

When ΔV is negative, the electrons lose kinetic energy and gain potential energy as they travel from the emitter to the collector; the change in potential energy of an electron is $\Delta U = q\ \Delta V = -e\ \Delta V$, a positive quantity when ΔV is negative. The most energetic electrons have kinetic energies $K_{\max}$. The stopping potential corresponds to the largest potential difference needed for these electrons to lose all their kinetic energy. Conservation of energy ($\Delta U = -\Delta K$) then gives

$$eV_0 = K_{\max}. \tag{45-9}$$

Curve b in Fig. 45-4 shows the result of a repetition of the stopping potential measurement in which we have doubled the intensity of the incident light. Comparison of curves a and b in the figure suggests the first of three important facts about the photoelectric effect:

> *Fact 1. If light of a given wavelength falls on a given emitter, the stopping potential does not depend on the intensity of the incident light.*

Now let us measure the stopping potential V_0 when light of *different* wavelengths falls on a given emitter. Figure 45-5,

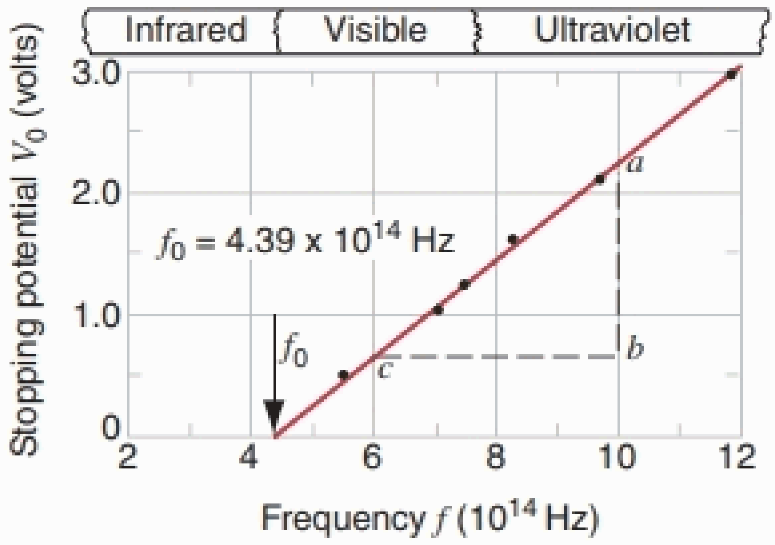


FIGURE 45-5. The stopping potential as a function of frequency for a sodium emitter. Note the cutoff frequency f_0. Photoelectrons will not be ejected if the frequency of the incident light is less than this value.

in which we plot the stopping potential against the frequency (rather than the wavelength) of the incident light, shows the result for an emitter made of sodium. The plot is a straight line with an intercept f_0 on the frequency axis, suggesting a second fact about the photoelectric effect:

> *Fact 2. The frequency of the light falling on a given emitter must be greater than a certain value f_0. Otherwise the photoelectric effect will not occur.*

This *cutoff frequency* f_0 depends only on the material of which the emitter is made and is totally independent of the intensity of the incident light.

A third fact about the photoelectric effect has been firmly established by separate experiments:

> *Fact 3. Photoelectrons are emitted without delay once the incident light reaches the surface of the emitter.*

Experiment shows that "without delay" means "within about 10^{-9} s."

Einstein's Analysis of the Photoelectric Effect

In 1905, Einstein showed that his photon concept could explain the above three facts about the photoelectric effect. He wrote the following equation, now called *Einstein's photoelectric equation:*

$$hf = \phi + K_{\text{max}}. \qquad (45\text{-}10)$$

This equation suggests that a single photon carries an energy hf into the emitter, where it is essentially transferred to a single electron. Part of this energy, called the *work function* ϕ of the material that forms the emitter, is used up in causing the electron to escape from the emitter; you can think of it as a sort of surface energy tax. The remaining energy ($hf - \phi$) is the maximum energy K_{max} that the electron can have after it has emerged from the surface of the emitter.

Consider how Eq. 45-10 is consistent with the three experimental facts of photoelectricity that we have outlined above:

Fact 1: This follows because, if we double the intensity of the incident light, we simply double the number of photon–electron interactions in the emitter. Twice as many electrons leave the emitter, but the kinetic energy of the emitted electrons, and thus the stopping potential, remains unchanged.

Fact 2: By combining Eqs. 45-9 and 45-10 we can write

$$V_0 = (h/e)f - (\phi/e). \qquad (45\text{-}11)$$

Both h and e are fundamental physical constants and ϕ is a constant for a given emitter material. Thus Eq. 45-4 predicts that a plot of V_0 against f for a given emitter should be a straight line, its slope being h/e. Because we know the value of the electronic charge e, we can use the measured slope from Fig. 45-5 to find a value for the Planck constant h:

$$\text{slope} = \frac{ab}{bc} = \frac{2.30\text{ V} - 0.68\text{ V}}{10 \times 10^{14}\text{ Hz} - 6.0 \times 10^{14}\text{ Hz}}$$
$$= 4.1 \times 10^{-15}\text{ V}\cdot\text{s}.$$

Because the slope is equal to h/e (see Eq. 45-11), we can find the Planck constant h by multiplying this ratio by the electronic charge e,

$$h = (h/e)(e) = (4.1 \times 10^{-15}\text{ V}\cdot\text{s})(1.6 \times 10^{-19}\text{ C})$$
$$= 6.6 \times 10^{-34}\text{ J}\cdot\text{s},$$

in agreement with the value determined by other, independent, methods. In 1921, Einstein was awarded the Nobel Prize "for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect."

However, it has been known since 1927 (22 years after Einstein advanced his photon hypothesis) that the photoelectric effect can be explained equally well without invoking the photon! In 1969, W. E. Lamb and M. O. Scully showed this explicitly by applying the laws of quantum physics to the atoms of the emitter but treating the incident light as a classical electromagnetic wave. Thus we cannot say that the photoelectric effect experiments show in a *convincing* way that photons exist, and we must look to other experiments for proof of their existence.

SAMPLE PROBLEM 45-2. An emitter of sodium of area $A = 1.5\text{ cm}^2$ is placed a distance $r = 25$ cm from a light source of power output $P_0 = 1.0$ W. The light is monochromatic with a frequency well above the cutoff frequency for sodium. The work function ϕ of sodium is 2.3 eV. At what maximum rate can photoelectrons be emitted from the sodium emitter?

Solution If the source radiates uniformly in all directions, which we assume, the intensity I of the light at a distance r is given by

$$I = \frac{P_0}{4\pi r^2} = \frac{1.0\text{ W}}{4\pi(0.25\text{ m})^2} = 1.27\text{ W/m}^2.$$

The maximum possible rate of photoelectric events occurs when all of the incident light energy is absorbed by the photoelectric process. At 2.3 eV per event, this rate is

$$R = \frac{IA}{\phi} = \frac{(1.27\ \mathrm{W/m^2})(1.5\ \mathrm{cm^2})(10^{-4}\ \mathrm{m^2/cm^2})}{(2.3\ \mathrm{eV/event})(1.60 \times 10^{-19}\ \mathrm{J/eV})}$$
$$= 5.2 \times 10^{14}\ \mathrm{events/s}.$$

Show that this corresponds to a photoelectric current of 83 μA. In practice this maximum rate will not be achieved.

SAMPLE PROBLEM 45-3. Find the work function of sodium from the data plotted in Fig. 45-5.

Solution The intercept of the straight line in Fig. 45-5 on the frequency axis is the cutoff frequency f_0. Putting $V_0 = 0$ and $f = f_0$ in Eq. 45-11 yields

$$\phi = hf_0 = (6.63 \times 10^{-34}\ \mathrm{J \cdot s})(4.39 \times 10^{14}\ \mathrm{Hz})$$
$$= 2.91 \times 10^{-19}\ \mathrm{J} = 1.82\ \mathrm{eV}.$$

We note from Eq. 45-11 that a determination of the Planck constant h involves only the slope of the straight line in Fig. 45-5 and a determination of the work function ϕ involves only the intercept.

45-4 THE COMPTON EFFECT

In 1923, Arthur Holly Compton, then at Washington University in St. Louis, discovered that when a beam of x rays is scattered from a target, the wavelengths of the scattered x rays are slightly greater than the wavelength of the incident beam.

Figure 45-6 shows the experimental arrangement Compton used to observe this *Compton effect,* as we now call it. A beam of x rays of wavelength λ falls on a carbon (graphite) target T. Compton observed the x rays that were scattered at various angles ϕ to the incident beam and measured the intensity and the wavelength of the scattered rays at several of these angles.

Figure 45-7 shows his experimental results for four different values of ϕ. We see that although the incident beam consists essentially of a single wavelength λ, the scattered x rays have intensity peaks at two wavelengths. One of them (λ) is the same as the incident wavelength, but the other (λ') is larger by an amount $\Delta\lambda$. This *Compton shift* $\Delta\lambda$ varies with the angle ϕ at which the scattered x rays are observed.

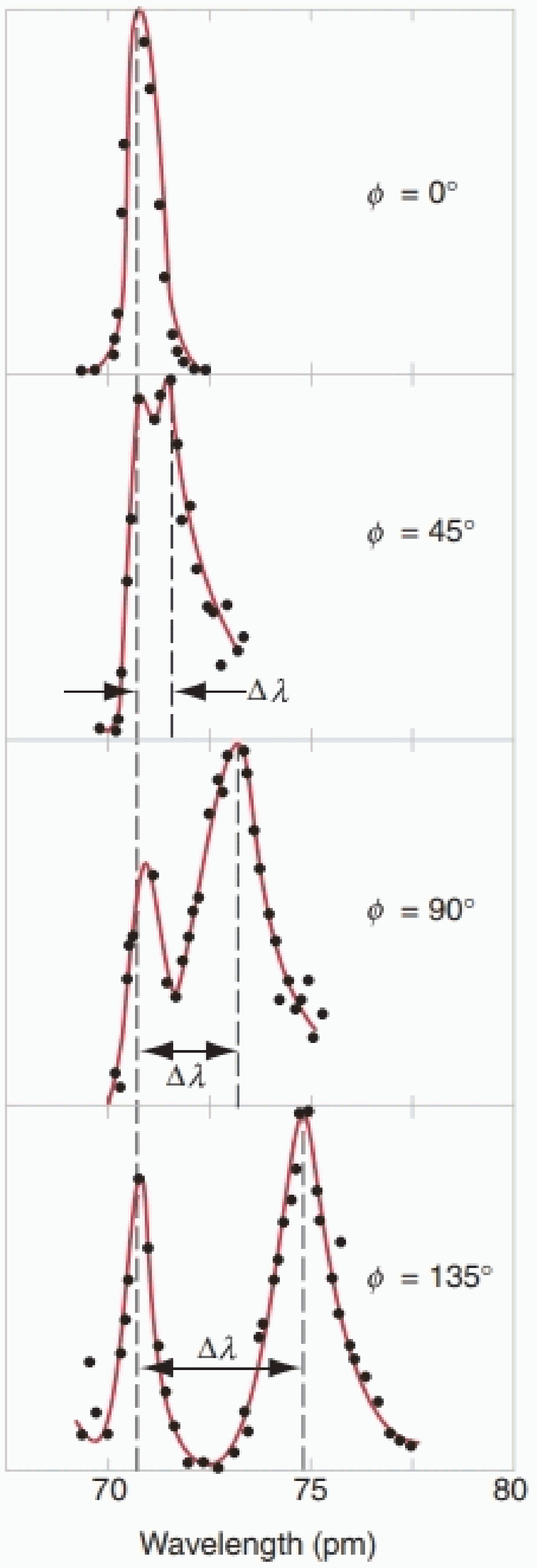


FIGURE 45-7. Compton's experimental results for four different values of the scattering angle ϕ. The peak at 71 pm is due to the scattering of the incident radiation (of wavelength 71 pm) from the tightly bound inner electrons of the atom. The second peak represents the radiation scattered from the loosely bound, nearly free outer electrons. For ϕ = 135°, the Compton shift $\Delta\lambda$ is about 4 pm.

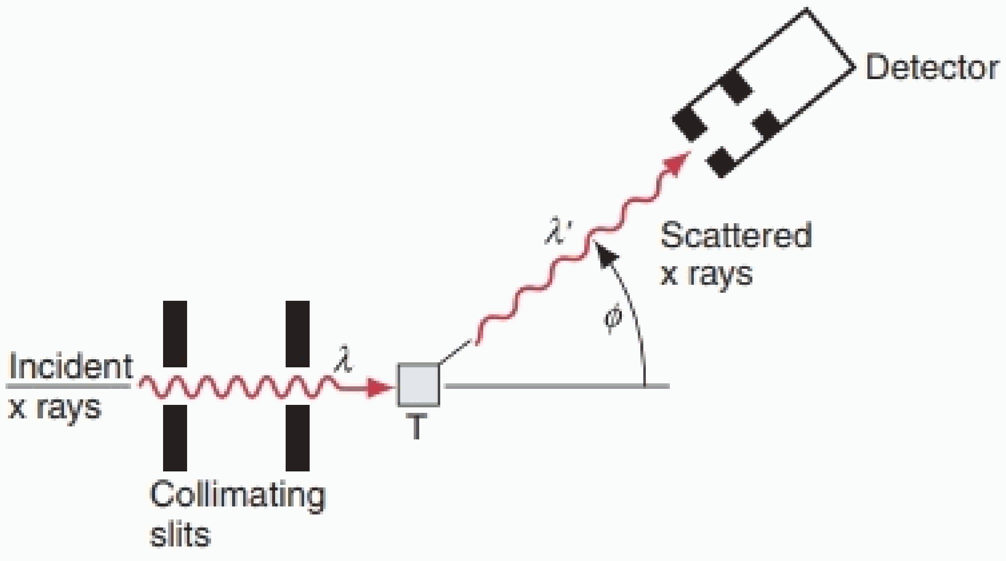


FIGURE 45-6. The experimental setup used for observing the Compton effect. The detector, which can be set to any desired angle ϕ, measured the intensity and the wavelength of the scattered x rays.

Compton derived an expression for $\Delta\lambda$ by analyzing the interaction between a photon of the incident x-ray beam and a free electron in the target as if it were a collision between two particles. First, let us analyze the situation qualitatively. Because the incident photon of energy E $(=hf)$ gives up some of its energy to the recoiling electron, the scattered photon must have a lower energy E' $(=hf')$. Because $E' < E$ we must have $f' < f$, which requires that $\lambda' > \lambda$. That is, the scattered x rays must have a larger wavelength than the incident x rays, exactly as Fig. 45-7 shows.

Now let us analyze the situation quantitatively. Figure 45-8 shows the initial and final states. We assume that the electron is initially at rest and is essentially free—that is, not bound to the atoms of the scatterer. This assumption holds well enough for the loosely bound outer electrons of the relatively light carbon atoms that make up the graphite target. Our plan is to find an expression for $\Delta\lambda$ by applying the laws of conservation of energy and of linear momentum to the interaction of Fig. 45-8. Because photons travel at the speed of light and the scattered electron may have a speed close to that of light, we must treat the analysis relativistically.

Conservation of energy for the photon–electron system requires that

$$E_{\text{initial}} = E_{\text{final}}$$

or

$$hf + mc^2 = hf' + mc^2 + K, \tag{45-12}$$

where mc^2 is the rest energy of the electron and K is the (relativistic) kinetic energy of the electron after the interaction. Substituting c/λ for f (and c/λ' for f') and using Eq. 20-27 for the relativistic kinetic energy, we have, as our energy conservation equation,

$$\frac{hc}{\lambda} = \frac{hc}{\lambda'} + mc^2\left[\frac{1}{\sqrt{1-(v/c)^2}} - 1\right]. \tag{45-13}$$

Now let us apply the (vector) law of conservation of linear momentum to the interaction of Fig. 45-8. The magnitudes of the linear momenta for the electron and the photon are, from Eqs. 20-23 and 45-2 respectively,

$$p_{\text{electron}} = \frac{mv}{\sqrt{1-(v/c)^2}} \quad \text{and} \quad p_{\text{photon}} = \frac{h}{\lambda}. \tag{45-14}$$

We can then write for the conservation of the x component of linear momentum

$$\frac{h}{\lambda} = \frac{h}{\lambda'}\cos\phi + \frac{mv}{\sqrt{1-(v/c)^2}}\cos\theta \tag{45-15}$$

and for the y component

$$0 = \frac{h}{\lambda'}\sin\phi + \frac{mv}{\sqrt{1-(v/c)^2}}\sin\theta. \tag{45-16}$$

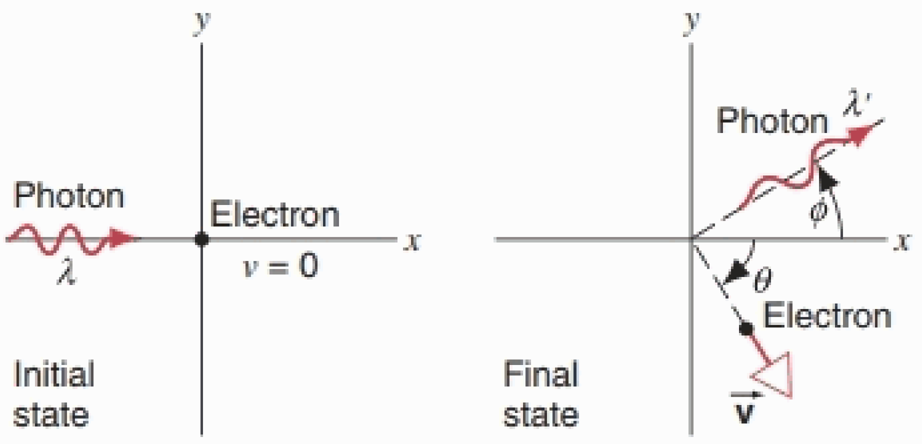


FIGURE 45-8. A Compton photon–electron interaction is treated as a collision between two particles. In the initial state, a photon of wavelength λ strikes a resting electron. In the final state, the photon recoils at an angle ϕ with an increased wavelength λ', and the electron recoils with speed v at an angle θ.

As defined in Fig. 45-8, the angle θ is negative, so that the second term of Eq. 45-16 is negative. Our aim is to use our three conservation equations (Eqs. 45-13, 45-15, and 45-16) to derive an expression for the Compton shift $\Delta\lambda$ $(=\lambda' - \lambda)$ so that we may compare it with Compton's experimental results.

Of the five variables (λ, λ', v, ϕ, and θ) that appear in our three conservation equations, we may eliminate two, leaving three unknowns. We choose to eliminate v and θ, which deal only with the electron (which Compton did not observe).

Carrying out the necessary algebraic steps (see Problem 10) leads to this result for the Compton shift:

$$\Delta\lambda = \lambda' - \lambda = \frac{h}{mc}(1 - \cos\phi). \tag{45-17}$$

Within experimental error, this equation agrees with the observed Compton shifts shown in Fig. 45-7.

Note in Eq. 45-17 that $\Delta\lambda$ does not depend on the wavelength λ of the incident radiation. Thus the *change* in wavelength $\Delta\lambda$ is the same for visible light ($\lambda \approx 500\text{ nm} = 5\times10^{-7}\text{ m}$) and for x rays ($\lambda \approx 70\text{ pm} = 7\times10^{-11}\text{ m}$). However, the *fractional change* in wavelength $\Delta\lambda/\lambda$ is much greater for x rays than for visible light, by a factor of about 70,000 for the examples cited. Thus the Compton shift is much easier to observe for x rays than for visible light.

It remains to explain the presence of the intensity peak in Fig. 45-7 for which the wavelength does not change on scattering. This peak results when an incoming photon interacts, not with a nearly free electron in the target, but with an electron tightly bound to an atom of the target. Because the entire target atom recoils during such an interaction, we must replace the electron mass m in Eq. 45-17 with M, the mass of a carbon atom. For carbon $M = 12.0$ u, which is about $22{,}000m$. If we replace m with M in Eq. 45-17, we see that, in agreement with experiment, the Compton shift for interactions with tightly bound electrons is negligible.

Does the Compton effect reveal the existence of the photon? Compton received a share of the 1927 Nobel Prize "for his discovery of the effect named after him." Recall that the photoelectric effect involved only the energy of the photon and not its momentum. The Compton effect, whose analysis involves both of these quantities, is even stronger evidence for the photon concept. However, since 1927 (four years after Compton performed his experiments) it has been shown that it is possible to derive the Compton shift equation (Eq. 45-17) without invoking the photon. Just as for the photoelectric effect, we can say that the Compton effect can be shown to be consistent with both the photon concept and the wave concept. We must look still further to find the experiment that can be understood *only* by the "light is a photon" hypothesis.

SAMPLE PROBLEM 45-4. X rays with $\lambda = 100$ pm are scattered from a carbon target. The scattered radiation is viewed at 90° to the incident beam. (*a*) What is the Compton shift $\Delta\lambda$? (*b*) What kinetic energy is imparted to the recoiling electron?

Solution (*a*) Putting $\phi = 90°$ in Eq. 45-17 we have, for the Compton shift,

$$\begin{aligned}\Delta\lambda &= \frac{h}{mc}(1 - \cos\phi)\\ &= \frac{6.63 \times 10^{-34}\ \text{J}\cdot\text{s}}{(9.11 \times 10^{-31}\ \text{kg})(3.00 \times 10^{8}\ \text{m/s})}(1 - \cos 90°)\\ &= 2.43 \times 10^{-12}\ \text{m} = 2.43\ \text{pm}.\end{aligned}$$

(*b*) Substituting c/λ for f (and c/λ' for f') in Eq. 45-12, we can write

$$\frac{hc}{\lambda} = \frac{hc}{\lambda'} + K.$$

Substituting $\lambda' = \lambda + \Delta\lambda$ and solving for K, we obtain

$$\begin{aligned}K &= \frac{hc\,\Delta\lambda}{\lambda(\lambda + \Delta\lambda)}\\ &= \frac{(6.63 \times 10^{-34}\ \text{J}\cdot\text{s})(3.00 \times 10^{8}\ \text{m/s})(2.43 \times 10^{-12}\ \text{m})}{(100 \times 10^{-12}\ \text{m})(100\ \text{pm} + 2.43\ \text{pm})(10^{-12}\ \text{m/pm})}\\ &= 4.72 \times 10^{-17}\ \text{J} = 295\ \text{eV}.\end{aligned}$$

You can show that the initial photon energy E in this case ($= hf = hc/\lambda$) is 12,400 eV, so the photon lost about 2.4% of its original energy in this interaction.

45-5 THE PHOTON REVEALED

Curiously enough, although the photon was proposed by Einstein in 1905, a strong case can be made that truly convincing experimental evidence for it emerged only in 1974 and especially in 1986, with the work that we describe in this section. What we seek is an experiment that not only supports the photon concept but that *cannot* also be explained by the wave concept.

Figure 45-9 suggests the broad outlines of a possible experiment. Light from source S falls on beam splitter B. A *beam splitter* is an optical device that splits a beam of light into two subbeams, a transmitted beam (beam X) and a reflected beam (beam Y), each of these subbeams having half the intensity of the incident beam. In the language of photons, a photon approaching a beam splitter has a 50% chance of being transmitted and a 50% chance of being reflected.

D_1 and D_2 in Fig. 45-9 are photoelectric light detectors that respond, respectively, to beams X and Y. The light source S is deliberately made feeble enough so that, on average, the total light energy in the apparatus at any given time is no greater than that of a single photon. Under these conditions each detector output will be a series of discrete electronic pulses.

If the photon model for light is correct, we expect that if a pulse appears at the output terminal of either detector at a given instant, no pulse will appear at the output terminal of the other detector at that same instant. Such *anticoincidences,* as they are called, can be detected and recorded electronically by arrangements not shown in Fig. 45-9.

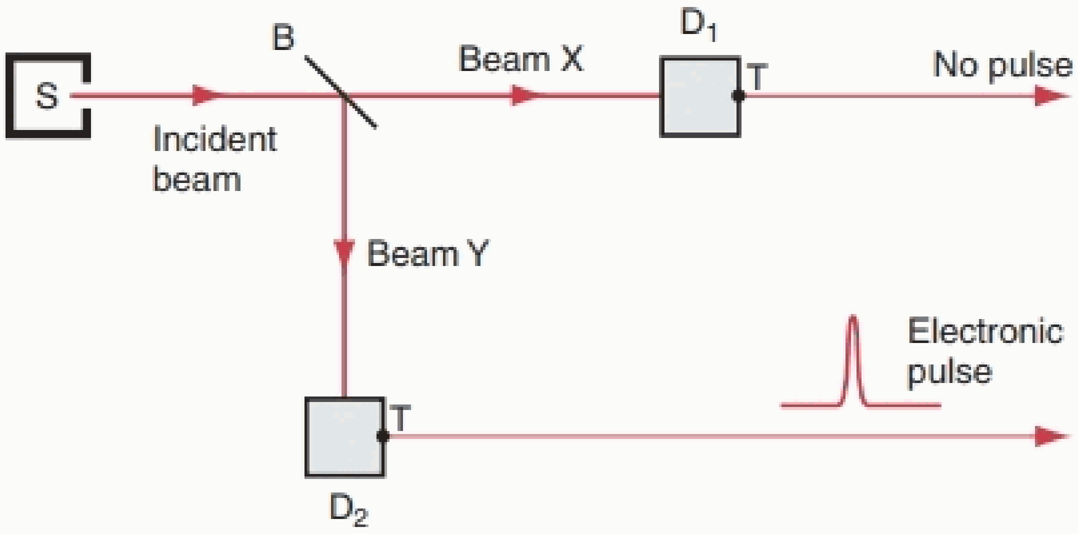


FIGURE 45-9. An anticoincidence apparatus. Light from source S falls on beam splitter B, where half is transmitted to detector D_1 and half is reflected to detector D_2. T marks the terminal at which the electronic output of each detector appears. To demonstrate photons, it should be true that if, at a certain time, an electronic pulse appears at the output of one detector (say, D_2) no pulse will appear at the output of the other detector (D_1). A typical anticoicidence is shown.

Early experiments of this kind showed no convincing pattern of anticoincidences and thus no confirmation of the photon model for light. Later it was shown that these negative results could be accounted for by small fluctuations in the intensity of the light source, among other factors. Thus the experiment, as we have described it so far, is not yet a convincing test for the presence or absence of photons.

Note that the apparatus of Fig. 45-9 cannot reveal the wave nature of light. Although the incident light beam is split into two subbeams—just as in the double-slit experiment—these beams enter separate detectors and are never recombined. Thus they cannot interfere with each other and cannot possibly produce the interference effects that are the hallmark of the wave model for light.

Using Designated Photons

In 1986, P. Grangier, G. Roger, and A. Aspect, following up the 1974 work of J. F. Clauser, modified the apparatus of Fig. 45-9 to eliminate its revealed shortcomings. They first arranged for detectors D_1 and D_2 to be inoperative (OFF) in their normal state. They could, however, turn the detectors ON at will—and for a brief interval—by applying an electronic "gate signal" of short duration to them.

The chief modification that these workers made, however, was to replace source S in Fig. 45-9 with a more elaborate arrangement. They caused a beam of calcium atoms to move through the source position and, by bombardment with laser light, caused some of these atoms to be raised to an excited state, labeled x in Fig. 45-10. In returning spontaneously to its ground state, labeled 0 in Fig. 45-10, each such calcium atom emitted two photons in a rapid cascade. For reasons that we will explain we label the first of these photons as the *trigger photon* and the other as the *designated photon.*

The rectangular box in Fig. 45-11 encloses the *designated photon source* used in this new experiment; it replaces

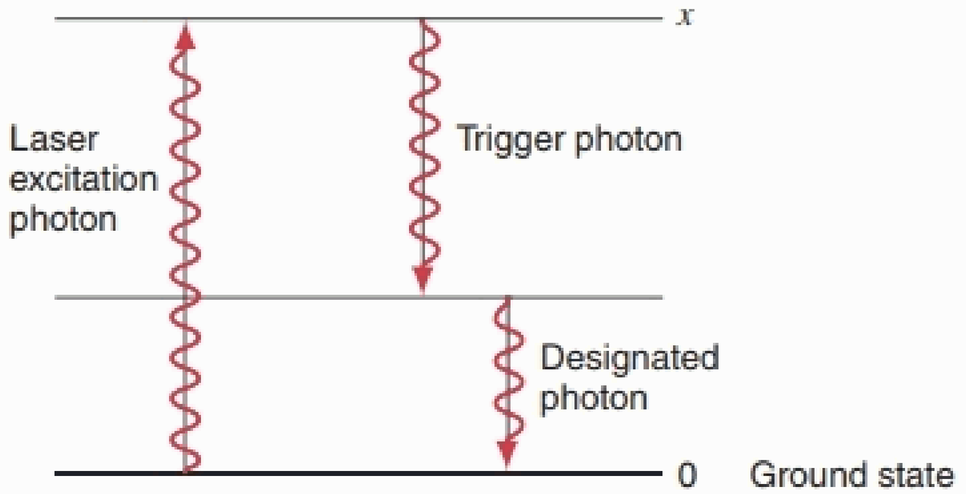


FIGURE 45-10. Three energy levels of the calcium atom are shown, the level marked 0 being the ground state. The atom is excited to state x by absorbing a photon from a laser beam. The atom then decays to its ground state by a two-step cascade process, with the two decay photons, labeled "trigger photon" and "designated photon," appearing essentially simultaneously.

source S in Fig. 45-9. In this new source, the trigger photon of Fig. 45-10 enters a detector G that generates an electronic gate signal at its terminal T. This signal, applied to detectors D_1 and D_2 of Fig. 45-9, turns them ON for a very short interval. Speaking very loosely, the trigger photon "tells" the detectors: "Wake up! A photon is coming. Shut down again after you have recorded it."

Note that the trigger photon is both emitted and absorbed within the source box and does not enter the apparatus of Fig. 45-9. The designated photon of Fig. 45-10, however, does enter this apparatus, moving toward beam splitter B. Arranging that the detectors are only ON when they need to be ensures that the detectors are not swamped with other photons that are unavoidably present.

With these modifications the experiment worked brilliantly. The expected anticoincidences were observed and the photon model for light was—for arguably the first time—established experimentally in a direct and convincing way. No wave interpretation of the results seems possible.

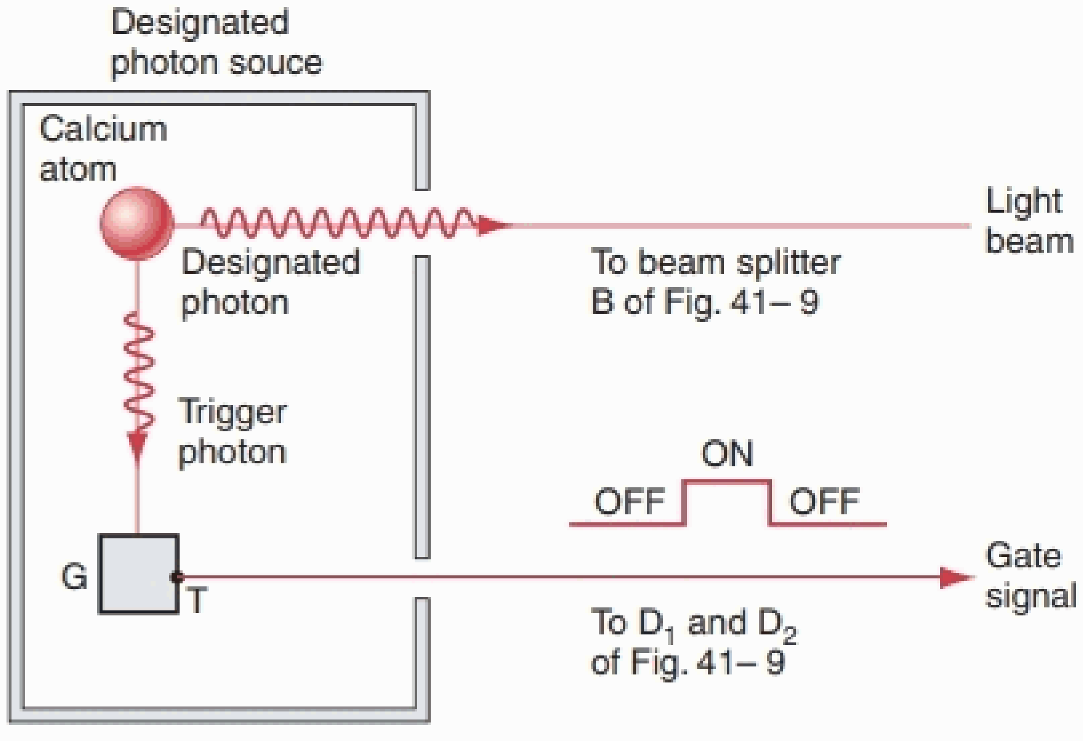


FIGURE 45-11. The apparatus in the rectangular box is a designated photon source, to be used in place of source S in Fig. 45-9. The calcium atom of Fig. 45-10 emits, in rapid succession, a trigger photon and a designated photon. The trigger photon is sent to an electronic gate generator G that sends an electronic gate signal to detectors D_1 and D_2 of Fig. 45-9, turning them briefly ON. The designated photon falls on the beam splitter of Fig. 45-9.

45-6 PHOTONS AND WAVES

We now have experimental evidence for two quite different models for light. "Light is a wave" is established by the double-slit experiment. "Light is a photon" is established by the anticoincidence experiment of Fig. 45-9. Figure 45-12 shows an extension of the apparatus of Fig. 45-9 with which, by making a simple modification, one can demonstrate at will either waves or photons. Experiments using this apparatus give us more insight into the ultimate nature of light.

Showing Waves

The light source S in Fig. 45-12 is the designated photon source of Fig. 45-11. The incident beam from this source is split at beam splitter B_1 into two subbeams, just as before. Beam X is then reflected from mirror M_1 and beam Y from mirror M_2. Mirror M_2 is fixed in place but mirror M_1 can be moved back and forth by a small controllable amount. Moving M_1 enables the experimenter to change, at will, the length of path X and thus to change the phase difference between the light following path X and that following path Y.

By tracing the beam paths in Fig. 45-12 you can see that beams X and Y (generated at beam splitter B_1) can be caused to recombine (by the action of beam splitter B_2) both at D_1 and also, separately, at D_2. The phase difference between the combined beams at either detector can be changed by moving mirror M_1. Thus the electronic outputs of both D_1 and D_2 should show a pattern of interference

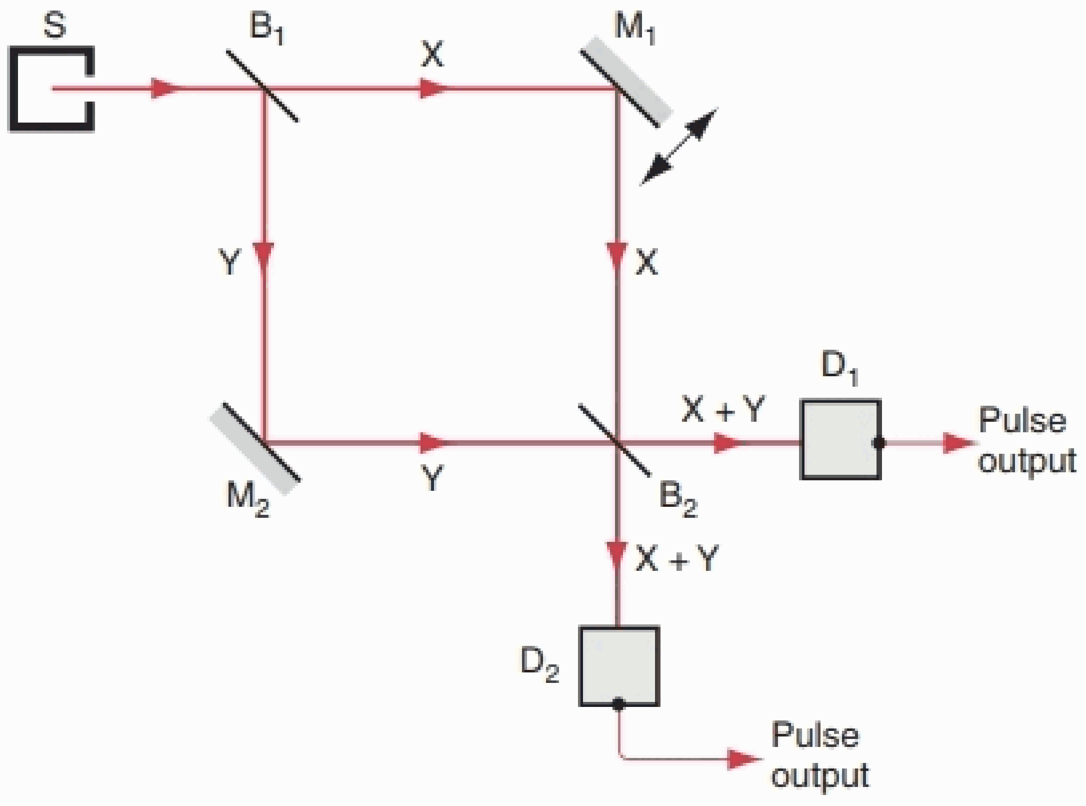


FIGURE 45-12. An extension of the apparatus of Fig. 45-9, designed to demonstrate both the wave and the photon properties of light. Test photons from source S follow two paths, labeled X and Y, through the apparatus. Light following each path is recombined at detector D_1 and also at detector D_2.

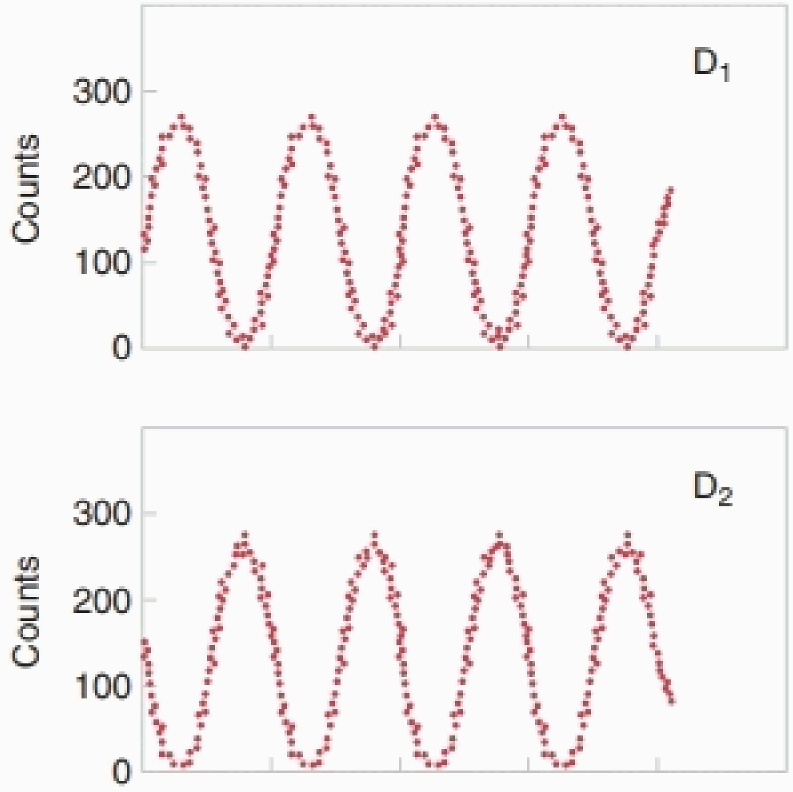


FIGURE 45-13. Counts accumulated in detectors D_1 and D_2 in Fig. 45-12 as a function of the position of movable mirror M_1. Each detector shows interference fringes, which we interpret by saying that light is wave-like in this experiment.

fringes as the experimenter moves this mirror. Figure 45-13 shows the pulse counting rates at the output of detectors D_1 and D_2 in Fig. 45-12 as a function of the position of mirror M_1. The interference patterns are clearly consistent with a wave nature for light.

In this mode of operation the apparatus of Fig. 45-12 is essentially a double-slit apparatus in that two subbeams are generated at beam splitter B_1, they follow paths of different lengths, and finally they meet again, both at detector D_1 and at detector D_2. It makes no difference that in the double-slit experiment the detector was a screen and in this case it is two separate electronic devices, each of which shows an interference pattern.

Showing Photons

If we remove beam splitter B_2, the apparatus of Fig. 45-12 is functionally identical to the apparatus of Fig. 45-9; that is, the two subbeams enter separate detectors and never recombine. Under these circumstances, the detector outputs will show a pattern of anticoincidences, consistent with a photon nature for light.

We conclude that if beam splitter B_2 is in place the experiment supports the wave model for light. If we remove B_2 the experiment supports the photon model. We are entitled to ask, "What is the nature of the light in transit through the apparatus of Fig. 45-12, *after* it leaves beam splitter B_1 but *before* it reaches beam splitter B_2?" Put another way: Does the light "know" whether or not beam splitter B_2 is in place before it reaches the location of B_2?

Delayed Choice Experiments

Clever *delayed choice experiments* can illuminate the questions we have just raised. An instructive experiment would be to leave beam splitter B_2 in place, as in Fig. 45-12. Then, just before the light reaches the location of B_2, rapidly snatch this beam splitter out of its path. Although it is not possible to remove and insert the beam splitter fast enough mechanically, the results of an equivalent experiment, involving rerouting light beams by electro-optical switches, was reported in 1987 by workers at the Max Planck Institute for Quantum Optics near Munich. We will report their results *as if* beam splitter B_2 were actually inserted and removed mechanically.

Suppose the experiment started with beam splitter B_2 in place. That is the setup for revealing interference fringes at the detector outputs and thus the wave nature of light. Can it be that the light in some way senses the presence of B_2 and therefore takes on its wave aspect? However, after the light has passed beam splitter B_1 but before it reaches the site of beam splitter B_2, we quickly remove that beam splitter from its path. Experiment shows that no interference is observed. If light "thought" that it was a wave after leaving B_1, it must have "changed its mind" when it reached the site of B_2 and did not find that beam splitter in place!

Conversely, let us start the experiment with B_2 removed. That is the setup for revealing anticoincidences at the detector outputs and thus the photon nature of light. Could it be that the light, having passed through B_1, in some way senses that B_2 is not there and takes on its photon aspect? Then, before the light reaches the site of B_2, we rapidly thrust this beam splitter across its path. Experiment shows that interference is always observed. If light "thought" that it was a photon after leaving B_1, it must have "changed its mind" when it reached the site of B_2 and "unexpectedly" found that beam splitter in place.

The experiment shows that you can insert beam splitter B_2 at any instant, from the start of the experiment to the very last picosecond, and the observed result at the detector outputs will be the same: Light is a wave. Conversely, you can remove B_2 at any time from the start of the experiment to the very last picosecond and the observed result at the detector outputs will be the same: Light is photons.

What are we to make of all this? Simply that, as the light is in transit through the apparatus, neither the wave nor the photon model—nor any combination of these classical concepts—makes any sense. It is simply "light" and, so far, no one has been able to propose a single acceptable "common sense" way of visualizing it. We can establish a model for light in the apparatus of Fig. 45-12 only when we confine our attention to the pulsed outputs of the two detectors. When beam splitter B_2 is permanently in place, the pulsed outputs show interference and we say "wave." When B_2 is permanently absent, the two outputs show anticoincidences and we say "photons." We can say absolutely nothing about the nature of light as it passes through the apparatus. All this reinforces our developing view that light is a subtle phenomenon that is not easily visualized or labeled. As we pointed out at the beginning of this chapter, light is neither exclusively particle-like nor exclusively wave-like.

Its particle and wave natures are intimately connected, and we can force it to reveal either type of behavior according to the kind of experiment we choose to perform.

45-7 SLOWING DOWN ATOMS BY PHOTON BOMBARDMENT

If you want to measure the properties of any object it is usually easiest to do so when the object is not moving. This is especially true if the object you wish to study is an atom. Unfortunately, as Sample Problem 45-5 shows, atoms of a gas at room temperature are far from at rest, typically moving with speeds exceeding that of a jet plane. In this section we wish to show how atoms can be slowed down from these high speeds by being bombarded with photons. This will, incidentally, provide further direct and convincing experimental evidence of the validity of the photon concept.

Figure 45-14 shows an atom of mass m in its ground state, illuminated by light of wavelength λ generated by a laser. Let the atom be moving with initial speed v_i in a direction opposite to that of the light that illuminates it. That is, the atom is moving upstream with respect to the laser beam. Suppose further that the atom absorbs a single photon of momentum (see Eq. 45-2) $p = h/\lambda$ from this beam. The photon disappears and its momentum is transferred entirely to the atom, reducing its upstream speed to v_f.

Conservation of momentum for the photon–atom interaction requires that

$$mv_i - h/\lambda = mv_f,$$

in which we have taken the direction of the moving atom as positive. Rearranging yields

$$h/\lambda = m(v_i - v_f) = -m\,\Delta v, \qquad (45\text{-}18)$$

where Δv, a negative quantity, is the change in speed of the atom after it has absorbed the photon. In Sample Problem 45-5 we calculate a typical value for Δv, finding it to be a few cm/s, about the speed of a running insect. Recall that v_i at room temperature is typically greater than that of a jet plane. Thus the absorption of a single photon can change the speed of an atom, but not by much. If we want to reduce the initial speed of the atom substantially we must illuminate the atom with a very intense laser beam so as to bombard it over and over again with photons.

When an atom absorbs a photon, the energy of the atom is raised from E_0, that of its ground state, to some higher level E_x, that of one of its excited states. Within a very short time (typically about 10^{-8} s) the atom will return spontaneously to its ground state, emitting a photon in the process. Indeed, if the atom did not return to its ground state it would not be in a position to absorb a second photon from the laser beam. The slowing down process for the atom requires a steady sequence of many photon absorption–emission events.

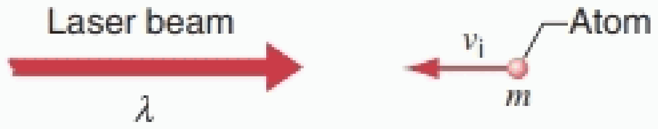


FIGURE 45-14. A beam of laser light of wavelength λ illuminates an atom of mass m that is moving upstream against the direction of the beam. The initial speed of the atom is v_i.

The photon emission associated with the return of the atom to its ground state will cause the atom to recoil. However, these recoils will occur in random directions so that, averaged over many photon absorption–emission events, the recoils do not greatly affect the magnitude of the directed motion of the atom although they may set a limit to the reduction in speed that can occur.

Suppose now that the atom were to move to the right in Fig. 45-14—that is, downstream with respect to the laser beam. Its speed would then be increased rather than decreased if it absorbed a photon. Convince yourself by studying Sample Problem 45-5 that, if an atom absorbs a photon, its change in speed Δv has the same magnitude no matter whether the atom is moving upstream or downstream. Speaking tentatively, it seems that if we bombard a group of atoms with a laser beam, half of them will slow down but the other half will speed up, resulting in no net change for the average speed of the group.

To see the clever way that experimenters have gotten around this difficulty, we must examine the photon absorption process more carefully. Although E_0, the ground-state energy of the atom, is sharply defined, the energy E_x of the excited state is not. Figure 45-15 is a plot of the relative probability $P(\lambda)$ that an atom—presumed to be at rest and in its ground state—will absorb a photon, raising the atom to a particular excited state of energy E_x. We see that there is a certain wavelength λ_x at which the absorption is most likely, but absorption can also occur at wavelengths either (very) slightly higher or (very) slightly lower than this value.

Figure 45-16 shows an atom illuminated by *two* laser beams, L and R, that point directly toward each other. For

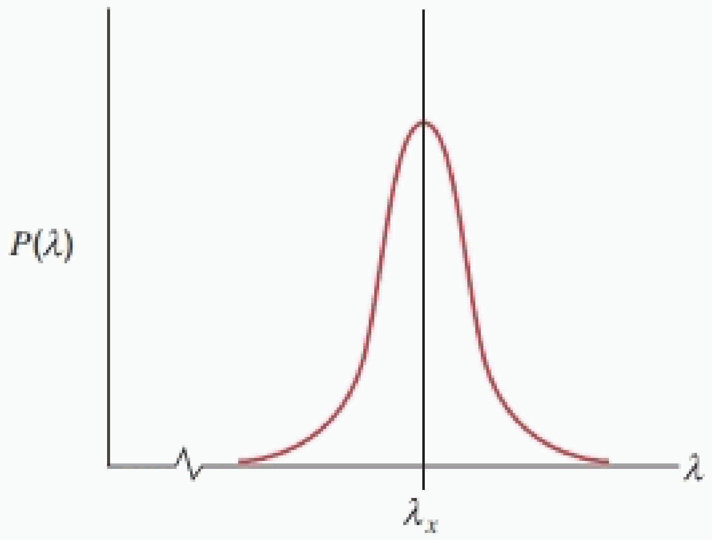


FIGURE 45-15. The relative probability $P(\lambda)$ that a resting atom in its ground state, illuminated by light of wavelength λ, will absorb a photon and move to a particular excited state of higher energy E_x. For light of a certain wavelength λ_x this probability is a maximum. However, a photon can still be absorbed, with reduced probability, for wavelengths either slightly higher or slightly lower than λ_x.

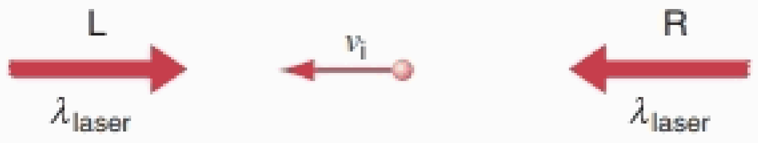


FIGURE 45-16. An atom in its ground state is illuminated by two beams of laser light, L and R, each of wavelength λ_{laser}. The atom is moving to the left, upstream with respect to beam L and downstream with respect to beam R.

reasons that we are about to explain the wavelength λ_{laser} of each laser beam is deliberately set—not to λ_x, but to a slightly larger value, as Fig. 45-17 shows.

The atom in Fig. 45-16 is moving to the left, upstream with respect to beam L and downstream with respect to beam R. From what we have learned so far, photons absorbed from beam L should slow the atom down while those absorbed from beam R should speed it up. These two effects do not cancel because of a consideration we have so far neglected to mention—namely, the Doppler effect (see Section 39-6).

Because of this effect the atom of Fig. 45-16, as it moves upstream in beam L, "senses" a higher light frequency—and thus a shorter wavelength—than it would if the atom were at rest. That is, the atom moving upstream "sees" a wavelength λ_L where $\lambda_L < \lambda_{\text{laser}}$ (see Sample Problem 45-6). A glance at Fig. 45-17 shows that the probability of photon absorption—and thus of slowing down—is considerably enhanced by this Doppler effect.

With respect to beam R, however, the atom is moving downstream and will "sense" a wavelength λ_R that is somewhat larger than λ_{laser}. Figure 45-17 now shows that the probability of photon absorption—and thus speeding up—is greatly reduced. By this clever arrangement an atom moving to the left is slowed down much more by beam L than it is speeded up by beam R, so there is a net reduction in speed. Convince yourself that an atom moving to the right in Fig. 45-16 will also experience a net reduction in speed.

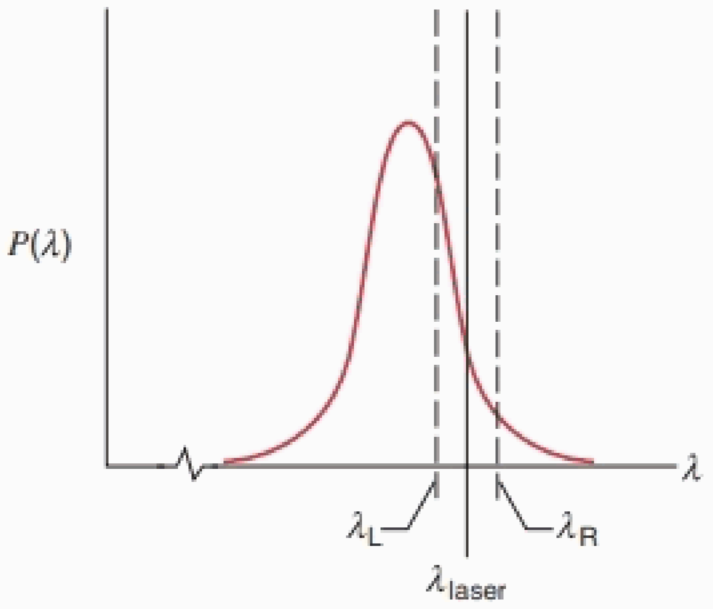


FIGURE 45-17. The atom of Fig. 45-16 is illuminated by laser light whose wavelength λ_{laser} is deliberately adjusted to be to the right of the maximum of the bell-shaped curve. As the atom moves to the left in Fig. 45-16, the Doppler effect will cause the atom to sense a shorter wavelength (λ_L) for light in beam L and a longer wavelength (λ_R) for light in beam R. Thus the probability that a photon will be absorbed from beam L is considerably enhanced and the probability that a photon will be absorbed from beam R is considerably reduced. If the atom is moving to the right in Fig. 45-16, the probability that a photon will be absorbed from beam R will be considerably enhanced and that from beam L considerably reduced. No matter in which direction the atom is moving, its speed will be reduced by the photon absorption mechanism.

Atoms of a gas do not move in one direction or the other along a single axis but are free to move in any direction. To take this into account, the experimenters illuminated a group of atoms with six laser beams, opposed in pairs and oriented in three mutually perpendicular directions. Thus, no matter what the direction of a moving atom, each of its velocity components is reduced in magnitude by the Doppler mechanism we have described. An atom that finds itself at the intersection of the six laser beams behaves like an insect trapped in a viscous fluid. No matter in which direction it moves it is slowed down. Experimenters say that atoms so confined are trapped in *optical molasses.* Because reducing the average speed of a group of atoms corresponds to reducing their temperature, the method we have described is called *laser cooling.*

Around 1985, Steven Chu and his coworkers at the Bell Laboratories succeeded in slowing down a group of about a million sodium atoms by laser cooling. To the naked eye, the groups of atoms seemed like a glowing cloud the size of a pea, located at the intersection of the six laser beams. Although the atoms were slowed down (to about 30 cm/s) they were not trapped. To hold them in place against their tendency to fall under gravity, the experimenters incorporated into their apparatus a trap that employed magnetic forces to counteract the downward force of gravity. With the atoms now both slowed and trapped, their properties could be studied at leisure.

Steven Chu, now at Stanford University, Claude Cohen-Tannoudji of the Collège de France and École Normale Supérieure, and William D. Phillips at the National Institute for Standards and Technology in Gaithersburg, Maryland, shared the 1997 Nobel Prize in physics "for development of methods to cool and trap atoms with laser light." The group led by Cohen-Tannoudji managed to reduce the temperature of a group of helium atoms to about 1.8×10^{-7} K ($= 0.18\ \mu$K) at which their mean speed was about 2 cm/s. In addition to fundamental studies in physics, the collective work of these three groups and others led to the development of the atomic clock at the National Institute of Standards and Technology (the cesium fountain clock; see Section 1-3) that is the present national frequency standard.

SAMPLE PROBLEM 45-5. (*a*) What is the root-mean-square speed v_{rms} of an atom of argon gas at room temperature ($T = 300$ K)? The mass m of an argon atom is 40.0 u. (*b*) An argon atom moving at its root-mean-square speed is traveling upstream into a laser beam whose wavelength is 105 nm. What is the change in speed of the atom if it absorbs a single photon from this beam?

Solution (*a*) Using Eq. 22-20, in which k is the Boltzmann constant, we have

$$v_{\mathrm{rms}} = \left(\frac{3kT}{m}\right)^{1/2} = \left[\frac{(3)(1.38 \times 10^{-23}\ \mathrm{J/K})(300\ \mathrm{K})}{(40.0\ \mathrm{u})(1.67 \times 10^{-27}\ \mathrm{kg/u})}\right]^{1/2}$$

$$= 430\ \mathrm{m/s}.$$

This is equal to 960 mi/h, which we can also express as Mach 1.3, considerably faster than a commercial jet plane. As Fig. 22-6 suggests, more than 40% of the argon atoms have speeds greater than this root-mean-square speed.

(*b*) Linear momentum must be conserved in the photon–atom interaction. Solving Eq. 45-18 for the speed reduction Δv yields

$$\Delta v = -\frac{h}{m\lambda} = -\frac{6.63 \times 10^{-34}\ \mathrm{J \cdot s}}{(40.0\ \mathrm{u})(1.67 \times 10^{-27}\ \mathrm{kg/u})(105 \times 10^{-9}\ \mathrm{m})}$$

$$= -9.45 \times 10^{-2}\ \mathrm{m/s} \approx -9\ \mathrm{cm/s}.$$

This is about the speed of a running insect. Recall that the initial speed of the atom was 430 m/s.

SAMPLE PROBLEM 45-6. A resting sodium atom can be raised from its ground state to an excited state if it absorbs a photon from a laser beam whose wavelength is 589 nm, presumed for simplicity to be exact. Because of the Doppler effect, what wavelength shift will the sodium atom "see" for that laser beam if it is moving upstream in the beam with a speed of 500 m/s?

Solution The atom is moving slowly with respect to the speed of light, so we can use Eq. 39-21 and neglect the terms beyond the linear one. Revised for upstream motion and expressed in terms of wavelength rather than frequency, this equation becomes

$$\lambda = \lambda_0 (1 + v/c)^{-1} \approx \lambda_0 (1 - v/c).$$

Here v is the speed of the atom and λ_0 is the wavelength of the laser light. As usual, c is the speed of light. We can rewrite this equation further, expressing it as a Doppler wavelength shift, or

$$\begin{aligned}\Delta\lambda &= \lambda_0 - \lambda = \lambda_0 (v/c)\\ &= (589\ \mathrm{nm})(500\ \mathrm{m/s})/(3.00 \times 10^{8}\ \mathrm{m/s})\\ &= 9.8 \times 10^{-4}\ \mathrm{nm}.\end{aligned}$$

Thus the moving atom will "see" the wavelength of the laser light increased by this small amount. The fractional change in wavelength because of the Doppler effect is $(9.8 \times 10^{-4}\ \mathrm{nm})/(589\ \mathrm{nm})$ or 1.7×10^{-6}; this amounts to only 0.00017%.

MULTIPLE CHOICE

45-1 Introducing the Photon

45-2 Thermal Radiation

1. The spectral radiancy of a cavity radiator is a maximum at a wavelength $\lambda_{\max} = \lambda_0$. The temperature of the body is now changed so that the spectral radiancy is a maximum at $\lambda_{\max} = \lambda_0/2$. This temperature change also causes the radiant intensity of the body to increase by a factor of
   (A) 2. (B) 4. (C) 8. (D) 16.

2. A cavity radiator has its maximum spectral radiancy at a wavelength of 28.98 μm, in the infrared region of the spectrum. The temperature of the body is now increased so that the radiant intensity $I(T)$ of the body is doubled.
   (*a*) What is this new temperature?
   (A) $\sqrt[4]{100}$ K (B) $\sqrt[4]{200}$ K
   (C) $\sqrt[4]{2}\ 100$ K (D) $\sqrt{4}\ 100$ K
   (*b*) At what wavelength will the spectral radiancy now have its maximum value?
   (A) $\sqrt[4]{28.98}\ \mu$m (B) $\sqrt[4]{28.98/2}\ \mu$m
   (C) $\sqrt[4]{1/2}\ 28.98\ \mu$m (D) $\sqrt{1/4}\ 28.98\ \mu$m

3. Which of these statements is a consequence of Planck's derivation of the radiation law?
   (A) Atomic oscillators can emit and absorb energy at discrete values only.
   (B) Atomic oscillators can emit and absorb energy at discrete frequencies only.
   (C) Both (A) and (B)
   (D) Neither (A) nor (B)

45-3 The Photoelectric Effect

4. What is the effect of increasing the wavelength of the light that falls on the emitter in a photoelectric-effect apparatus?
   (A) The work function decreases.
   (B) The cutoff frequency decreases.
   (C) The stopping potential decreases.
   (D) The time delay for emission of photoelectrons increases.
   (E) None of the above

5. What is the effect of increasing the intensity of the light that falls on the emitter in a photoelectric-effect apparatus?
   (A) The work function decreases.
   (B) The cutoff frequency decreases.
   (C) The stopping potential decreases.
   (D) The time delay for emission of photoelectrons increases.
   (E) None of the above

6. Monochromatic light with a frequency well above the cutoff frequency is incident on the emitter in a photoelectric-effect apparatus. The frequency of the light is then doubled while the intensity is kept constant.
   (*a*) How does this affect the stopping potential?
   (A) The stopping potential will increase.
   (B) The stopping potential will decrease.
   (C) The stopping potential will remain the same.
   (*b*) How does this affect the photoelectric current?
   (A) The photoelectric current will increase.
   (B) The photoelectric current will decrease.
   (C) The photoelectric current will remain the same.

45-4 The Compton Effect

7. Compton scattering describes a collision between a photon and an electron. Which of the following is false?
   (A) Compton scattering conserves momentum.
   (B) Compton scattering is elastic.

(C) Compton scattering is proof of the particle nature of light.

(D) Equation 45-17 assumes that the electron was originally at rest.

8. How does the maximum Compton shift $\Delta\lambda_{\max}$ depend on the incident wavelength λ?

(A) $\Delta\lambda_{\max} \propto \lambda^2$ (B) $\Delta\lambda_{\max} \propto \lambda$

(C) $\Delta\lambda_{\max}$ is independent of λ.

(D) $\Delta\lambda_{\max} \propto 1/\lambda$

9. A photon of wavelength λ enters an electron gas. What is the minimum number of collisions that could result in the photon being completely absorbed in the gas?

(A) $\approx \lambda mc/2h$ (B) $\approx \lambda mc/h$

(C) $\approx 2\lambda mc/h$ (D) $\approx \lambda mc^2/h$

10. Photon *A* has twice the energy of photon *B*. What is the ratio of the momentum of *A* to that of *B*?

(A) 4 : 1 (B) 2 : 1 (C) 1 : 1 (D) 1 : 2

45-5 The Photon Revealed

11. Consider early anticoincidence experiments. Which of the following could be a reasonable explanation for why both detectors would output a pulse?

(A) Two photons were simultaneously emitted from the source.

(B) Two photons were emitted at different times from the source, but the detectors have some inherent time uncertainties.

(C) Light can behave like a wave with an arbitrarily small amplitude.

(D) All of the above are possible explanations.

12. Which of the following is most significant about the 1986 modifications in the anticoincidence experiments?

(A) With two photons emitted the signals would be twice as prominent.

(B) Using excited calcium atoms helped to insure that photons were emitted only one at a time.

(C) The trigger photon helped to insure that only one photon would be present in the detector regions at any one time.

(D) Using laser light resulted in only coherent waves being emitted.

45-6 Photons and Waves

13. A very weak light source, which emits only one photon at a time, is allowed to emit light through a double slit. The light then shines on a photographic film.

(*a*) What will be seen on the film if it is exposed for only a very short period of time?

(A) The film will show only discrete spots where photons struck the film, with no evidence of an interference pattern.

(B) The film would show discrete spots where photons struck the film, but the spots would be distributed similar to an interference pattern.

(C) The film would show a continuous interference pattern.

(*b*) Instead the film is exposed for a very long period of time. What is seen on the film now?

(A) The film will show only discrete spots where photons struck the film, with no evidence of an interference pattern.

(B) The film would show discrete spots where photons struck the film, but the spots would be distributed similar to an interference pattern.

(C) The film would show a continuous interference pattern.

45-7 Slowing Down Atoms by Photon Bombardment

14. Which of the following is essential for laser cooling to take place? Choose all that apply.

(A) Atoms can absorb photons that have energies only within a narrow range.

(B) The light source must be coherent.

(C) Light must obey the Doppler effect.

QUESTIONS

1. "Pockets" formed by the coals in a coal fire seem brighter than the coals themselves. Is the temperature in such a pocket appreciably higher than the surface temperature of an exposed glowing coal? Explain this common observation.

2. The relation $I(T) = \sigma T^4$ (Eq. 45-3) is exact for true cavities and holds for all temperatures. Why do we not use this relation as the basis of a definition of temperature at, say, 100°C?

3. Do all incandescent solids obey the fourth-power law of temperature, as Eq. 45-3 seems to suggest?

4. By simply looking at the sky at night, can you pick out stars that are hotter than the Sun? Cooler than the Sun? What do you look for? Is the star's brightness a clue?

5. Betelgeuse, the prominent red star in the constellation Orion, has a surface temperature that is much lower than that of the Sun, yet it radiates energy into space at a much higher rate than the Sun does. How can that be?

6. Less than a few percent of the energy supplied to a 100-W lamp appears in the form of visible light. What happens to the rest of it? What could be done to increase this percentage? Why has it not already been done?

7. Your skin temperature is about 300 K. In what region of the electromagnetic spectrum do you emit thermal radiation most intensely?

8. Spectral radiancy curves for cavity radiators at different temperatures do not intersect; see Fig. 45-1. Suppose, however, that they did. Can you show that this would violate the second law of thermodynamics?

9. We claim that all objects radiate energy by virtue of their temperature and yet we cannot see all objects in the dark. Why not?

10. Show that the Planck constant has the dimensions of angular momentum. Does this necessarily mean that angular momentum is a quantized quantity?

11. For quantum effects to be "everyday" phenomena in our lives, what order of magnitude value would *h* need to have? [See G. Gamow, *Mr. Tompkins in Wonderland* (Cambridge University

Press, Cambridge, 1957), for a delightful popularization of a world in which the physical constants c, G, and h make themselves obvious.]

12. Explain why a tube used to examine photoelectric emission is (*a*) evacuated and (*b*) fitted with a window made of quartz rather than glass.

13. Determine whether or not relativistic mechanics is needed to verify the photoelectric equation (Eq. 45-10) with a 1% uncertainty. Note that typical stopping potentials are a few volts.

14. In Fig. 45-4, why does the photoelectric current not rise vertically to its maximum (saturation) value when the applied potential difference is slightly more positive than V_0?

15. In the photoelectric effect, why does the existence of a cutoff frequency speak in favor of the photon theory and against the wave theory?

16. Why are photoelectric measurements so sensitive to the nature of the photoelectric surface?

17. An insulated metal plate yields photoelectrons when first illuminated by ultraviolet light but then does not give up any more. Explain.

18. Why is it that even for incident radiation that is monochromatic the photoelectrons are emitted with a spread of velocities?

19. We claim that all the energy of an absorbed photon is given to an emitted photoelectron. Why can we neglect the energy taken by the lattice?

20. Is the Compton effect more supportive of the photon theory of light than the photoelectric effect? Explain your answer.

21. Consider the following procedures: (*a*) bombard a metal with electrons, (*b*) place a strong electric field near a metal, (*c*) illuminate a metal with light, (*d*) heat a metal to a high temperature. Which of these procedures can result in the emission of electrons?

22. A certain metal plate is illuminated by light of a definite frequency. Whether or not photoelectrons are emitted as a result depends on which of the following features: (*a*) intensity of illumination, (*b*) length of time of exposure to the light, (*c*) thermal conductivity of the plate, (*d*) area of the plate, (*e*) material of the plate?

23. Does Einstein's theory of photoelectricity, in which light is postulated to be a stream of photons, invalidate the double-slit interference experiment in which light is postulated to be a wave?

24. Explain the statement that one's eyes could not detect faint starlight if light were not particle-like.

25. How can a photon energy be given by $E = hf$ when the very presence of the frequency f in the formula implies that light is a wave?

26. A photon has no rest mass since it can never be at rest with respect to any observer. If energy equals mc^2, how can a photon have any energy?

27. The momentum p of a photon is given by $p = h/\lambda$. Why is it that c, the speed of light, does not appear in this expression?

28. In discussing the propagation of light we sometimes use straight rays, sometimes waves, and at still other times discrete photons. To what extent, if at all, are these views compatible with one another? Are there cases in which one view is clearly superior to the others?

29. Given that $E = hf$ for a photon, the Doppler shift in frequency of radiation from a receding light source would seem to indicate a reduced energy for the emitted photons. Is this in fact true? If so, what happened to the conservation of energy principle? (See "Questions Students Ask," *The Physics Teacher,* December 1983, p. 616.)

30. How does a photon differ from a material particle?

31. What is the direction of a Compton scattered electron with maximum kinetic energy compared with the direction of the incident monochromatic photon beam?

32. Why, in the Compton scattering picture (Fig. 45-8), would you expect $\Delta\lambda$ to be independent of the materials of which the scatterer is composed?

33. Why do we not observe a Compton effect with visible light?

34. Light from distant stars is Compton scattered many times by free electrons in outer space before reaching us. This shifts the light toward the red. How can this shift be distinguished from the Doppler red shift due to the motion of receding stars?

35. In both the photoelectric effect and the Compton effect there is an incident photon and an ejected electron. What is the difference between these two effects?

36. List and discuss the assumptions made by Planck in connection with the cavity radiation problem, by Einstein in connection with the photoelectric effect and by Compton in connection with the Compton effect.

37. Describe several experimental methods that can be used to determine the value of the Planck constant h.

38. A laser projects a beam of light across a laboratory table. If you put a diffraction grating in the path of the beam and observe the spectrum, you declare the beam to be a wave. If instead you put a clean metal surface in the path of the beam and observe the ejected photoelectrons, you declare this same beam to be a stream of particles (photons). What can you say about the beam if you do not put anything in its path?

EXERCISES

45-1 Introducing the Photon

1. (*a*) By using the "best" values of the fundamental constants, as found in Appendix B, show that the energy E of a photon is related to its wavelength λ by

$$E = \frac{1240\ \mathrm{eV\cdot nm}}{\lambda}.$$

This result can be useful in solving many problems. (*b*) The orange-colored light from a highway sodium lamp has a wavelength of 589 nm. How much energy is possessed by an individual photon from such a lamp?

2. Consider monochromatic light falling on a photographic film. The incident photons will be recorded if they have enough energy to dissociate a AgBr molecule in the film. The minimum energy required to do this is about 0.60 eV. Find the cutoff wavelength greater than which the light will not be recorded. In what region of the spectrum does this wavelength fall?

3. An atom absorbs a photon having a wavelength of 375 nm and immediately emits another photon having a wavelength of 580 nm. What was the net energy absorbed by the atom in this process?

4. Under ideal conditions the normal human eye will record a visual sensation at 540 nm if incident photons are absorbed at a rate as low as 100 s^{-1}. To what power level does this correspond?

5. Most gaseous ionization processes require energy changes of 1.0×10^{-18} to 1.0×10^{-16} J. What region then of the Sun's electromagnetic spectrum is chiefly responsible for creating the ionosphere in the Earth's atmosphere?

6. In the photon picture of radiation, show that if two parallel beams of light of different wavelengths are to have the same intensity, then the rates per unit area at which photons pass through any cross section of the beams are in the same ratio as the wavelengths.

7. An ultraviolet lightbulb, emitting at 400 nm, and an infrared lightbulb, emitting at 700 nm, each are rated at 130 W. (*a*) Which bulb radiates photons at the greater rate? (*b*) How many more photons does it generate per second than the other bulb?

8. A special kind of lightbulb emits monochromatic light at a wavelength of 630 nm. It is rated at 70.0 W and is 93.2% efficient in converting electrical energy to light. How many photons will the bulb emit over its 730-h lifetime?

45-2 Thermal Radiation

9. In 1983, the Infrared Astronomical Satellite (IRAS) detected a cloud of solid particles surrounding the star Vega, radiating maximally at a wavelength of 32 pm. What is the temperature of this cloud of particles? Assume an emissivity of unity.

10. Low-temperature physicists would not consider a temperature of 2.0 mK (0.0020 K) to be particularly low. At what wavelength is the spectral radiancy of a cavity at this temperature a maximum? To what region of the electromagnetic spectrum does this radiation belong? What are some of the practical difficulties of operating a cavity radiator at such a low temperature?

11. Calculate the wavelength of maximum spectral radiancy and identify the region of the electromagnetic spectrum to which it belongs for each of the following: (*a*) the 2.7-K cosmic background radiation, a remnant of the primordial fireball; (*b*) your body, assuming a skin temperature of 34°C; (*c*) a tungsten lamp filament at 1800 K; (*d*) the Sun, at an assumed surface temperature of 5800 K; (*e*) an exploding thermonuclear device, at an assumed fireball temperature of 10^7 K; (*f*) the universe immediately after the big bang, at an assumed temperature of 10^{38} K. Assume cavity radiation conditions throughout.

12. (*a*) The effective surface temperature of the Sun is 5800 K. At what wavelength would you expect the Sun to radiate most strongly? In what region of the spectrum is this? Why then does the Sun appear yellow? (*b*) At what temperature is cavity radiation most visible to the human eye? See Fig. 39-6.

13. A cavity whose walls are held at 1900 K has a small hole, 1.00 mm in diameter, drilled in its wall. At what rate does energy escape through this hole from the cavity interior?

14. A cavity at absolute temperature T_1 radiates energy at a power level of 12.0 mW. At what power level does the same cavity radiate at temperature $2T_1$?

15. (*a*) An ideal radiator has a spectral radiancy at 400 nm that is 3.50 times its spectral radiancy at 200 nm. What is its temperature? (*b*) What would be its temperature if its spectral radiancy at 200 nm were 3.50 times its spectral radiancy at 400 nm?

45-3 The Photoelectric Effect

16. Satellites and spacecraft in orbit about the Earth can become charged due, in part, to the loss of electrons caused by the photoelectric effect induced by sunlight on the space vehicle's outer surface. Suppose that a satellite is coated with platinum, a metal with one of the largest work functions: $\phi = 5.32$ eV. Find the smallest-frequency photon that can eject a photoelectron from platinum. (Satellites must be designed to minimize such charging.)

17. You wish to pick a substance for a photocell operable with visible light. Which of the following will do (work function in parentheses): tantalum (4.2 eV), tungsten (4.5 eV), aluminum (4.2 eV), barium (2.5 eV), lithium (2.3 eV), cesium (1.9 eV)?

18. Find the maximum kinetic energy in eV of photoelectrons if the work function of the material is 2.33 eV and the frequency of the radiation is 3.19×10^{15} Hz.

19. (*a*) The energy needed to remove an electron from metallic sodium is 2.28 eV. Does sodium show a photoelectric effect for red light, with $\lambda = 678$ nm? (*b*) What is the cutoff wavelength for photoelectric emission from sodium and to what color does this wavelength correspond?

20. Light of wavelength 200 nm falls on an aluminum surface. In aluminum, 4.2 eV is required to remove an electron. What is the kinetic energy of (*a*) the fastest and (*b*) the slowest emitted photoelectrons? (*c*) Find the stopping potential. (*d*) Calculate the cutoff wavelength for aluminum.

21. Incident photons strike a sodium surface having a work function of 2.28 eV, causing photoelectric emission. When a stopping potential of 4.92 V is imposed, there is no photocurrent. Find the wavelength of the incident photons.

22. The stopping potential for photoelectrons emitted from a surface illuminated by light of wavelength 491 nm is 710 mV. When the incident wavelength is changed to a new value, the stopping potential is found to be 1.43 V. (*a*) What is this new wavelength? (*b*) What is the work function for the surface?

23. (*a*) If the work function for a metal is 1.85 eV, what would be the stopping potential for light having a wavelength of 410 nm? (*b*) What would be the maximum speed of the emitted photoelectrons at the metal's surface?

24. A lithium surface for which the work function is 2.49 eV is irradiated with light of frequency 6.33×10^{14} Hz. The loss of electrons causes the metal to acquire a positive potential. What must this potential have become by the time its value prevents further loss of electrons from the surface?

25. The data in Fig. 45-5 were obtained by Robert A. Millikan (United States, 1868–1953), who was awarded the 1921 Nobel Prize in physics for his verification of the photoelectric effect. Millikan's photoelectric data for lithium are:

Wavelength (nm)	433.9	404.7	365.0	312.5	253.5
Stopping potential (V)	0.55	0.73	1.09	1.67	2.57

Make a plot like Fig. 45-5, which is for sodium, and find (*a*) the Planck constant and (*b*) the work function for lithium.

26. To remove an inner, most tightly bound, electron from an atom of molybdenum requires an energy of 20 keV. If this is to be done by allowing a photon to strike the atom, (*a*) what must be the associated wavelength of the photon? (*b*) In what region of the spectrum does the photon lie? (*c*) Could this process be called a photoelectric effect? Discuss your answers.

45-4 The Compton Effect

27. A particular x-ray photon has a wavelength of 41.6 pm. Calculate the photon's (*a*) energy, (*b*) frequency, and (*c*) momentum.

28. Find (*a*) the frequency, (*b*) the wavelength, and (*c*) the momentum of a photon whose energy equals the rest energy of the electron.

29. By how much does a sodium atom slow down on absorbing a photon of wavelength 589 nm with which it collides head-on?

30. The quantity h/mc in Eq. 45-17 is often called the *Compton wavelength*, λ_C, of the scattering particle and that equation is written

$$\Delta\lambda = \lambda_C(1 - \cos\phi).$$

(*a*) Calculate the Compton wavelength of an electron. Of a proton. (*b*) What is the energy of a photon whose wavelength is equal to the Compton wavelength of the electron? Of the proton? (*c*) Show that in general the energy of a photon whose wavelength is equal to the Compton wavelength of a particle is the same as the rest energy of that particle.

31. Photons of wavelength 2.17 pm are incident on free electrons. (*a*) Find the wavelength of a photon that is scattered 35.0° from the incident direction. (*b*) Do the same if the scattering angle is 115°.

32. A 511-keV gamma-ray photon is Compton-scattered from a free electron in an aluminum block. (*a*) What is the wavelength of the incident photon? (*b*) What is the wavelength of the scattered photon? (*c*) What is the energy of the scattered photon? Assume a scattering angle of 72.0°.

33. Show that $\Delta E/E$, the fractional loss of energy of a photon during a Compton collision, is given by

$$\frac{\Delta E}{E} = \frac{hf'}{mc^2}(1 - \cos\phi).$$

34. What fractional increase in wavelength leads to a 75% loss of photon energy in a Compton collision with a free electron?

35. Find the maximum wavelength shift for a Compton collision between a photon and a free proton.

36. A 6.2-keV x-ray photon falling on a carbon block is scattered by a Compton collision and its frequency is shifted by 0.010%. (*a*) Through what angle is the photon scattered? (*b*) How much kinetic energy is imparted to the electron?

37. An x-ray photon of wavelength $\lambda = 9.77$ pm is backscattered by an electron ($\phi = 180°$). Determine (*a*) the change in wavelength of the photon, (*b*) the change in energy of the photon, and (*c*) the final kinetic energy of the electron.

38. Calculate the fractional change in photon energy for a Compton collision with ϕ in Fig. 45-8 equal to 90° for radiation in (*a*) the microwave range, with $\lambda = 3.00$ cm; (*b*) the visible range, with $\lambda = 500$ nm; (*c*) the x-ray range, with $\lambda = 0.10$ nm; and (*d*) the gamma-ray range, with $\lambda = 1.30$ pm. What are your conclusions about the importance of the Compton effect in these various regions of the electromagnetic spectrum, judged solely by the criterion of energy loss in a single Compton encounter?

39. Through what angle must a 215-keV photon be scattered by a free electron so that it loses 10.0% of its energy?

45-5 The Photon Revealed

40. A certain photon detector can resolve the time of arrival of a photon to within 10^{-8} s. Two of these detectors are used for an anticoincidence experiment. (*a*) What is the maximum average rate of photon emission from a source if there is any hope of demonstrating anticoincidence? (*b*) Assuming the source emits light of wavelength 550 nm, what is the power output from the source?

45-6 Photons and Waves

45-7 Slowing Down Atoms by Photon Bombardment

41. Unexcited sodium atoms at rest will absorb photons from a light source whose wavelength is 588.995 nm. The sodium atom is originally traveling at a speed 300 m/s "upstream" into a laser beam. (*a*) What should be the wavelength of the photons in the laser beam that can be absorbed by this sodium atom? (*b*) What will be the speed of the sodium atom immediately after it absorbs a photon? (*c*) The excited sodium atom then emits another photon in the direction of its velocity. What is the final speed of the atom?

42. Argon atoms slow down by about 15 cm/s for each photon emitted and absorbed. (*a*) If the argon atom is originally traveling at 430 m/s, how many photons must the atom interact with in order to slow the speed to 50 cm/s? (*b*) Assuming the cooling apparatus is about 1 meter in length, at what rate must the atom interact with photons to cool before it escapes from the apparatus?

PROBLEMS

1. An oven with an inside temperature $T_0 = 215°\text{C}$ is in a room with a temperature of $T_1 = 26.2°\text{C}$. There is a small opening of area $A = 5.20\ \text{cm}^2$ in one side of the oven. How much net power is transferred from the oven to the room?

2. The filament of a particular 100-W lightbulb is a cylindrical wire of tungsten 0.280 mm in diameter and 1.80 cm long. See Appendix D for needed data on tungsten. Assume that the filament radiates identically to a cavity radiator and ignore absorption of energy by the filament from the surroundings. (*a*) Calculate the operating temperature of the filament. (*b*) How long does it take for the filament to cool by 500 C° after the bulb is switched off?

3. A convex lens 3.8 cm in diameter and of focal length 26 cm produces an image of the Sun on a thin black screen the same size as the image. Find the highest temperature to which the screen can be raised. The effective temperature of the Sun is 5800 K.

4. Show that the wavelength $\lambda_{\max}$ at which Planck's spectral radiation law, Eq. 45-6, has its maximum is given by Eq. 45-4,

$$\lambda_{\max} = (2898\ \mu\text{m}\cdot\text{K})/T.$$

(Hint: Set $dR/d\lambda = 0$; an equation will be encountered whose numerical solution is 4.965.)

5. Consider a planet, with radius R, revolving about the Sun in a circular orbit of radius r. Suppose that the planet has no atmosphere (and therefore no "greenhouse effect" on its surface temperature). (*a*) Show that the surface temperature T of the planet is given from the relation $T^4 = P_{\text{Sun}}/16\pi\sigma r^2$ where P_{Sun} is the radiant power output of the Sun. (*b*) Evaluate the temperature numerically for the Earth.

6. (*a*) By integrating the Planck radiation law, Eq. 45-6, over all wavelengths, show that the power radiated per square meter of a cavity surface is given by

$$I(T) = \left(\frac{2\pi^5 k^4}{15h^3c^2}\right)T^4 = \sigma T^4.$$

(Hint: Make a change in variables, letting $x = hc/\lambda kT$. The definite integral

$$\int_0^\infty \frac{x^3\,dx}{e^x - 1}$$

will be encountered, which has the value $\pi^4/15$.) (*b*) Verify that the numerical value of the constant σ is $5.67 \times 10^{-8}\ \text{W/m}^2\cdot\text{K}^4$.

7. Assume that a 100-W sodium-vapor lamp radiates its energy uniformly in all directions in the form of photons with an associated wavelength of 589 nm. (*a*) At what rate are photons emitted from the lamp? (*b*) At what distance from the lamp will the average flux of photons be 1.00 photon/(cm$^2\cdot$s)? (*c*) At what distance from the lamp will the average density of photons be 1.00 photon/cm^3? (*d*) Calculate the photon flux and the photon density 2.00 m from the lamp.

8. Show, by analyzing a collision between a photon and a free electron (using relativistic mechanics), that it is impossible for a photon to give all its energy to the free electron. In other words, the photoelectric effect cannot occur for completely free electrons; the electrons must be bound in a solid or in an atom.

9. X rays with a wavelength of 71.0 pm eject photoelectrons from a gold foil, the electrons originating from deep within the gold atoms. The ejected electrons move in circular paths of radius r in a region of uniform magnetic field $\vec{\mathbf{B}}$. Experiment shows that $rB = 188\ \mu\text{T}\cdot\text{m}$. Find (*a*) the maximum kinetic energy of the photoelectrons and (*b*) the work done in removing the electrons from the gold atoms that make up the foil.

10. Carry out the necessary algebra to eliminate v and θ from Eqs. 45-13, 45-15, and 45-16 to obtain the Compton shift relation, Eq. 45-17.

11. (*a*) Show that when a photon of energy E scatters from a free electron, the maximum recoil kinetic energy of the electron is given by

$$K_{\max} = \frac{E^2}{E + mc^2/2}.$$

(*b*) Find the maximum kinetic energy of the Compton-scattered electrons knocked out of a thin copper foil by an incident beam of 17.5-keV x rays.

CHAPTER 46

THE NATURE OF MATTER

In Chapter 45 we learned that light, long regarded as a wave, has an equally convincing particle aspect—namely, the photon. In this chapter we present experimental evidence supporting the claim that matter, long regarded as made up of particles, has an equally convincing wave aspect. Heisenberg's uncertainty principle will show us the limit to which we can extend the concept of "particle" into the quantum realm.

We then introduce Schrödinger's equation, the fundamental equation of quantum mechanics, which deals with the wave behavior of particles. Among the predictions of quantum mechanics is the phenomenon of barrier penetration, in which particles can enter regions of space that would be forbidden to them in the classical realm of Newtonian mechanics.

46-1 MATTER WAVES

Nature abounds in symmetries of all kinds. If light can be both wave and particle, is it possible that matter can also have these two aspects? Can the electron, traditionally regarded as a particle since its discovery in 1898, also have a wave aspect? More specifically, can we assign a wavelength and a frequency to a moving electron?

In 1924, the French physicist Prince Louis-Victor de Broglie (1892–1987), motivated entirely by this symmetry argument, answered yes to these questions. He proposed that an electron of energy E and linear momentum p can be described by a *matter wave* whose wavelength and frequency are given by

$$\lambda = \frac{h}{p} \tag{46-1}$$

and

$$f = \frac{E}{h}, \tag{46-2}$$

in which h is the Planck constant. Note that these equations are exactly the same as those that we used to describe the photon in Section 45-1, except that here we have solved them for λ and f instead of for p and E. The wavelength of a moving particle calculated using Eq. 46-1 is called its *de Broglie wavelength.* De Broglie shared the 1929 Nobel Prize for his discovery of the wave nature of matter.

Objects such as marbles or baseballs do not seem at all wave-like to us. We can understand why this is so because the Planck constant h in Eq. 46-1 is so small and the momentum p of even slowly moving macroscopic objects is so large that the calculated de Broglie wavelengths of such objects are small indeed, being many orders of magnitude smaller than the size of an atomic nucleus. It seems impossible to design an experiment to reveal the wave nature of such large-scale objects. In the next section, we will see that this wave-like behavior is quite common for electrons and other microscopic particles.

SAMPLE PROBLEM 46-1. Calculate the de Broglie wavelength of (*a*) a dust particle of mass 1.0×10^{-9} kg drifting with a speed of 2.0 cm/s, and (*b*) an electron whose kinetic energy is 120 eV.

Solution (*a*) Using Eq. 46-1, we find

$$\lambda = \frac{h}{p} = \frac{h}{mv}$$

$$= \frac{6.63 \times 10^{-34}\ \text{J}\cdot\text{s}}{(1.0 \times 10^{-9}\ \text{kg})(2.0 \times 10^{-2}\ \text{m/s})} = 3.3 \times 10^{-23}\ \text{m}.$$

This is smaller than the size of a typical nucleus by a factor of about 50 million. Imagine how small must be the de Broglie wavelength of a baseball traveling at 90 mi/h. No apparatus can be built that would reveal the wave nature of such moving objects. (*b*) Because $K \ll mc^2$, we may safely ignore relativity and use the classical relationship between momentum and kinetic energy; namely, $K = p^2/2m$. From Eq. 46-1, then

$$\begin{aligned}\lambda &= \frac{h}{p} = \frac{h}{\sqrt{2mK}}\\ &= \frac{6.63 \times 10^{-34}\ \mathrm{J \cdot s}}{\sqrt{(2)(9.11 \times 10^{-31}\ \mathrm{kg})(120\ \mathrm{eV})(1.6 \times 10^{-19}\ \mathrm{J/eV})}}\\ &= 1.1 \times 10^{-10}\ \mathrm{m} = 0.11\ \mathrm{nm}.\end{aligned}$$

The electron's de Broglie wavelength, which is larger than that for the dust particle of part (*a*) by a factor of about 3×10^{12}, is about the size of an atom. We will show in the following section how wavelengths of this magnitude can be measured, thus verifying that electrons do indeed possess a wave-like aspect.

46-2 TESTING DE BROGLIE'S HYPOTHESIS

The straightforward way to demonstrate the wave nature of a moving electron is to allow a beam of such particles to fall on a screen containing a double slit and look for interference fringes as the electrons merge from the slits. This is how we demonstrated the wave nature of light. As Sample Problem 46-1 shows, the de Broglie wavelength of even a relatively low-momentum electron ($K = 120$ eV) is only about the size of a single atom. The technology in de Broglie's day did not permit the construction of slits as closely spaced as this.

However, the de Broglie wavelength calculated in Sample Problem 46-1 (0.11 nm) is comparable to the spacing between atoms in a typical crystalline solid such as copper or aluminum. The atoms that form such a solid can serve as a sort of three-dimensional diffraction grating to reveal the wave nature of an incident beam of electrons. Experiments based on this principle were successfully carried out in 1927, by C. J. Davisson and L. H. Germer at what was then the Western Electric Laboratories (now the Bell Laboratories of Lucent Technologies) and by George P. Thomson and A. Reid at the University of Aberdeen in Scotland. The results of each of these experiments showed that moving electrons can indeed be assigned a wavelength calculated using Eq. 46-1. Davisson and Thomson shared the 1937 Nobel Prize for their work. We now describe several experiments that support the hypothesis that electrons have a wave-like aspect.

G. P. Thomson's Experiment. Figure 46-1*a* shows an arrangement similar to that used by Thomson and Reid to demonstrate the wave nature of the electron. A monoenergetic beam of electrons is allowed to fall on a target in the form of a fine powder of tiny, randomly oriented metallic crystallites. In such a target, it is virtually certain that, by chance, the incident electron beam will always encounter a small number of crystallites so oriented that Eq. 43-12 is satisfied and a so-called Bragg reflection (see Section 43-4) of the incident electron beam will occur. From symmetry, electrons undergoing these Bragg reflections will form a number of concentric circles on a photographic film placed as shown in Fig. 46-1*a*. Each circle corresponds to a particular set of atomic planes in the target material.

Figure 46-1*b* shows the results of a modern repetition of Thomson's experiment, using finely powdered aluminum as a target. Figure 46-1*c* shows the pattern that is formed if the electron beam is replaced with a beam of x rays. (The energy of the x-ray photons and the momentum of the electrons were chosen so that both beams had the same wavelength.) A simple glance at these two diffraction patterns leaves no doubt that both originated in the same way. If x rays are wave-like in this experiment (they are) then so are electrons.

The Davisson–Germer Experiment. The apparatus used by Davisson and Germer is illustrated in Fig. 46-2*a*. A beam of electrons from a heated filament is accelerated through a potential difference ΔV of 54 V and falls on a highly polished single crystal of nickel. The lines of atoms on the surface form a sort of diffraction grating for the elec-

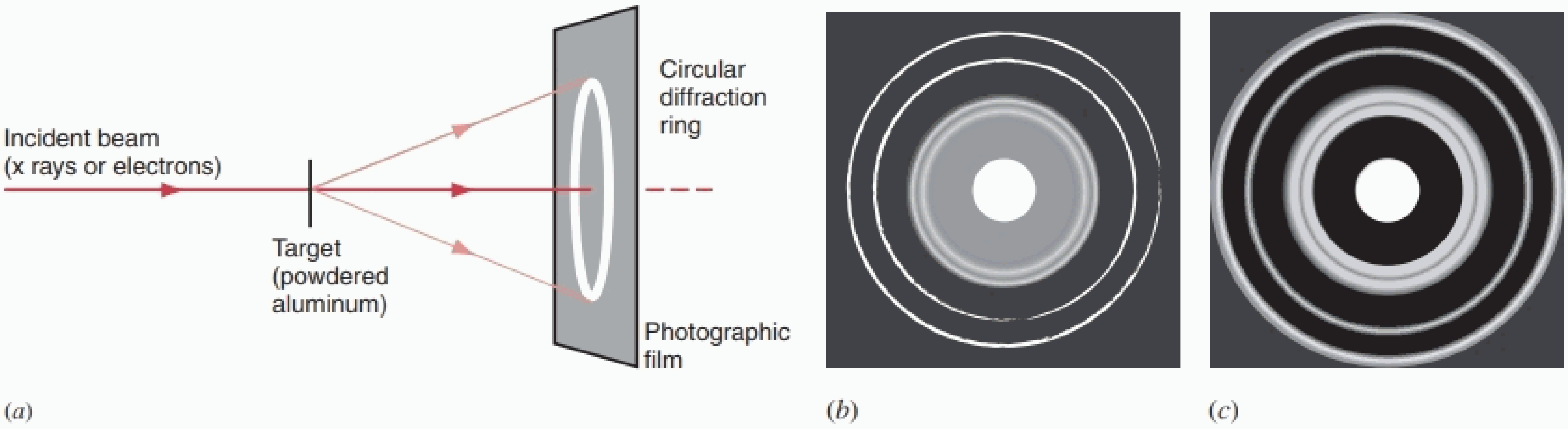


FIGURE 46-1. (*a*) An arrangement to demonstrate the wave nature of the electrons in a beam incident from the left. (*b*) The diffraction pattern formed when an electron beam falls on a target of powdered aluminum. (*c*) The pattern formed when a beam of x rays is substituted for the electron beam.

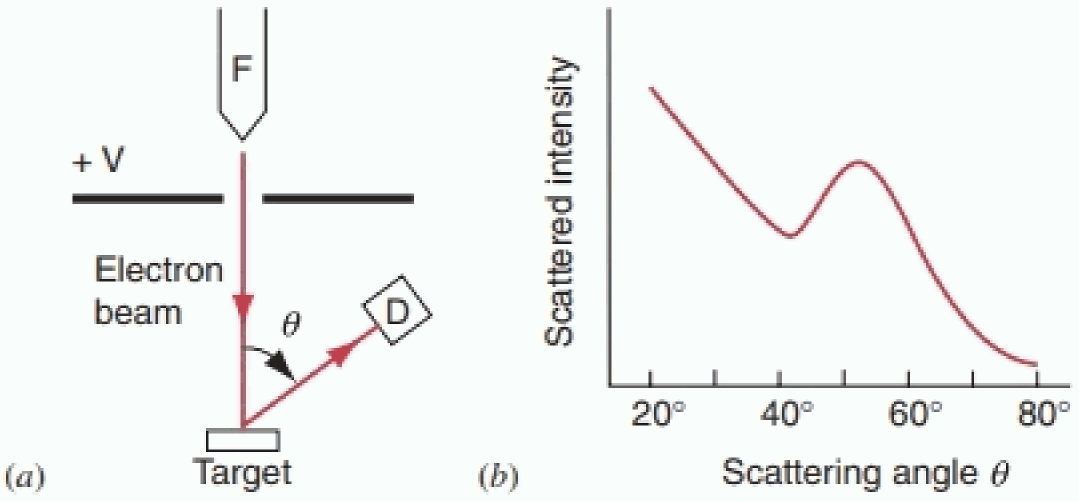


FIGURE 46-2. (*a*) The apparatus used in the Davisson-Germer experiment. A beam of electrons from the heated filament F is accelerated through a voltage V and strikes a target consisting of a single crystal of nickel. The scattered electrons are observed in detector D which can be moved to various angles θ relative to the incident beam. (*b*) For $V = 54$ V, the detector intensity shows an interference maximum at $\theta = 52°$.

tron waves with a spacing equal to the lattice spacing of the nickel crystal, or 0.215 nm. A detector can be moved to any angle θ relative to the direction of the original beam of electrons.

The geometry of Fig. 46-2*a* is identical to that commonly used with reflection-type optical diffraction gratings to measure the wavelength of light. The grating equation $d \sin \theta = m\lambda$ (Eq. 43-1) also applies in the geometry of crystal diffraction. As shown in Fig. 46-2*b*, Davisson and Germer observed a first-order ($m = 1$) interference maximum at about $\theta = 52°$ when their accelerating voltage was 54 V. For a 54-eV electron, the de Broglie wavelength is 0.167 nm. With $d = 0.215$ nm, the grating equation gives $\theta = \sin^{-1} (\lambda/d) = 51°$, in excellent agreement with their observations. This experiment not only provided striking confirmation of de Broglie's hypothesis, it also demonstrated the usefulness of low-energy electron beams as a probe of the properties of crystals, a technique that is still in common use today by materials scientists.

Diffraction by a Straight Edge. By early 1940s, techniques had advanced so that it became possible to observe the diffraction pattern formed by a beam of electrons that is allowed to fall, not on a single slit, but on the sharp straight edge of an opaque screen. When a beam of light is allowed to fall on such a straight edge, some light energy is diffracted a small distance into the geometrical shadow of the screen and—most prominently—a pattern of interference fringes appears outside the geometrical shadow, parallel to the edge. Figure 46-3*a* shows this pattern for light. Figure 46-3*b* shows a similar pattern for a beam of 38-keV electrons. (The electron's de Broglie wavelength was much smaller than that of the beam of visible light, so it was necessary to amplify the electron pattern by a factor of about 180,000, using an electron microscope, to bring the two figures to the same scale.) Once again, simple comparison of the two patterns of Fig. 46-3 leaves no doubt that they originated in the same way. If visible light behaves like a wave in this experiment—and it does—then so do electrons.

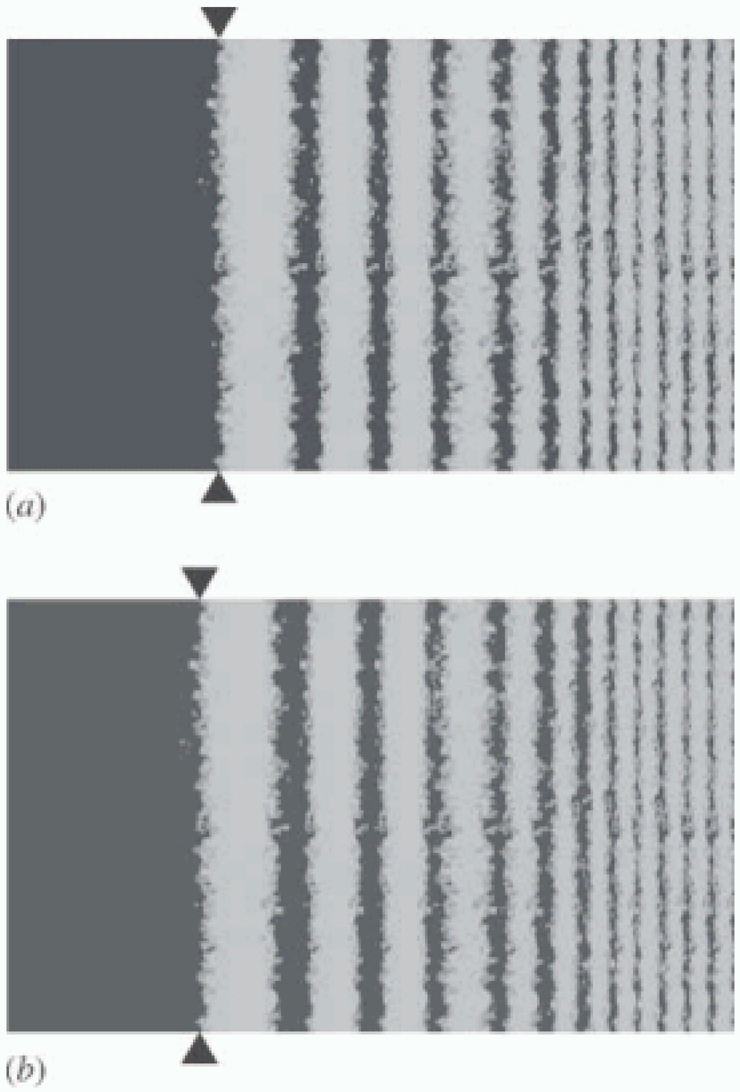


FIGURE 46-3. (*a*) A diffraction pattern formed when a beam of visible light falls on a straight edge. (*b*) The pattern formed when a beam of 38-keV electrons falls on a straight edge. Because the wavelengths in the two cases are quite different, the pattern in (*b*) has been amplified by a factor of about 180,000 to bring it to the same scale as the pattern in (*a*). In each case the small triangles represent the position of the straight edge, the geometrical shadow of the straight edge extending to the left.

Double-Slit Interference. As techniques continued to improve, it became possible to verify the wave nature of electrons by the traditional double-slit method, the method first used by Thomas Young in 1801 to establish the wave nature of light. Figure 46-4 shows a cross section of the double-slit arrangement used in 1989 by A. Tonomura and his colleagues at the Hitachi Advanced Research Laboratory and Gakushuin University in Tokyo. It consists of a thin, positively charged wire centered between two metallic plates that

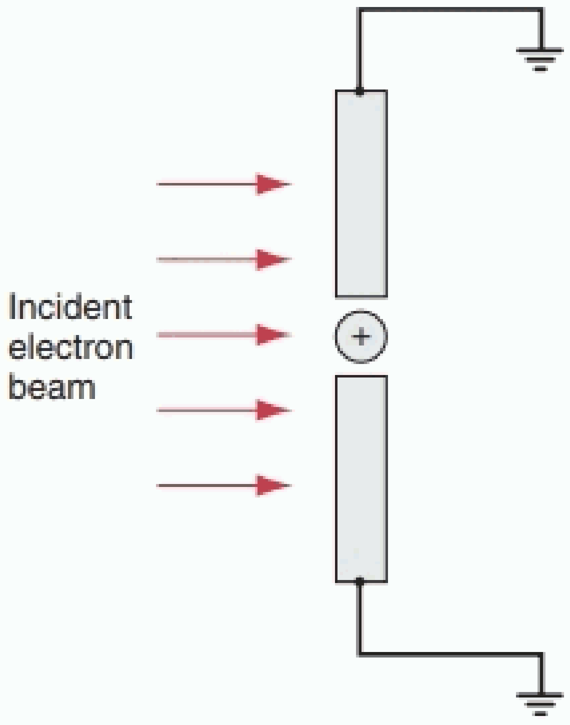


FIGURE 46-4. A double slit arrangement for electrons. Deflecting electric fields are set up in each slit by a thin, positively charged wire placed between two metal sheets held at ground potential.

are electrically grounded. Electrons from the incident beam were deflected by the electric force that acted on them as they passed through the slits. The emerging electrons—which had been accelerated to an energy of 50 keV—fell on a fluorescent screen, each electron marking its landing site by emitting, as a single pulse, a few hundred photons. After suitable amplification, it was arranged so that each of these feeble light pulses generated a visible dot on an image screen.

Figure 46-5 shows the results of four runs of increasing duration, in which increasingly large numbers of electrons were collected on the image screen. In Fig. 46-5*a* in which only about 100 electrons were collected, the distribution of image dots seems to be random. However, as the number of collected electrons increases—to 70,000 in Fig. 46-5*d*—a pattern of interference fringes slowly emerges: the classical demonstration of wave-like behavior. Quantitative analysis of the fringe spacing totally verifies Eq. 46-1.

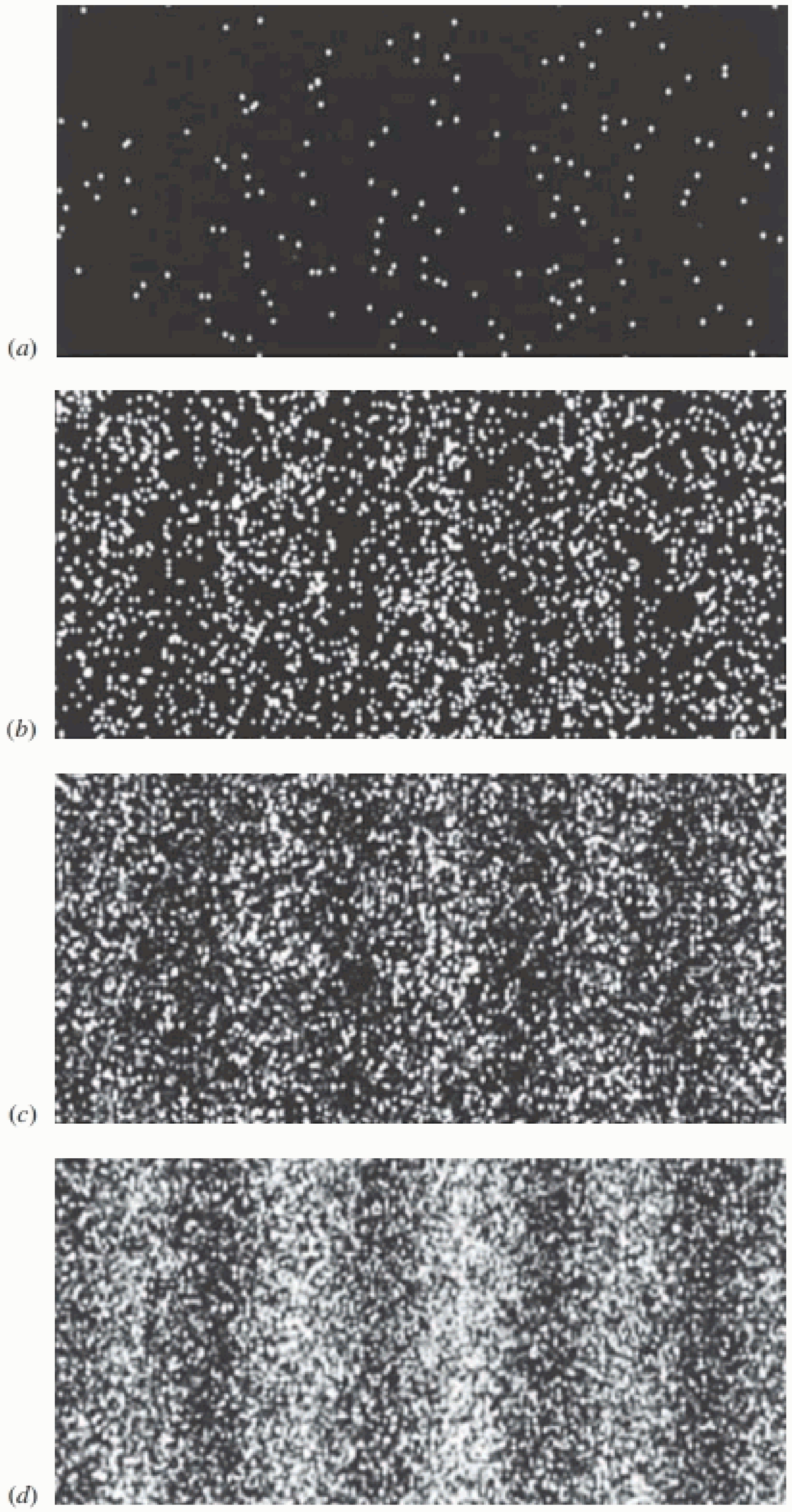


FIGURE 46-5. Four runs of the double-slit experiment of Tonomura and his associates. The numbers of electrons collected in each run are about (*a*) 100, (*b*) 3000, (*c*) 20,000, and (*d*) 70,000. Note how the interference fringes build up as more and more electrons are collected.

The Wave Nature of Other Particles

Equation 46-1 holds for electrons, as the experiments discussed above clearly demonstrate. Only additional experiments can tell whether it holds for other particles. Let us check it out for neutrons, using the double-slit method.

The double slit consists of a wire made of a neutron-absorbing material such as boron centered in a gap in a screen made of a similar neutron-absorbing material. Neutrons from a nuclear reactor are allowed to fall on this double-slit and a neutron detector is scanned across the neutron beam emerging on the far side of the slits in order to measure the beam intensity as a function of detector position. Figure 46-6 shows the results of one experiment, reported in 1991 by R. Gähler and A. Zeilinger. The dots are the experimental data points and the solid curve is the prediction of quantum mechanics. The excellent agreement between experiment and theory is solid confirmation of the wave nature of a neutron beam.

Figure 46-7 shows the interference fringes that resulted in a 1991 experiment in which a beam of helium atoms was allowed to fall on a screen containing a double slit. Interference fringes have also been demonstrated for beams of other atoms and for molecules of iodine vapor (I_2).

Diffraction studies by beams of electrons or neutrons are used routinely to study the atomic structures of solids or liquids. Such matter waves serve as a valuable supplement to x rays as tools for structure analysis. Electrons, for example, are less penetrating than x rays and so are particularly useful for studying surfaces. X rays interact largely with the electrons in a target sample, and for that reason it is not easy to use them to locate light atoms—particularly

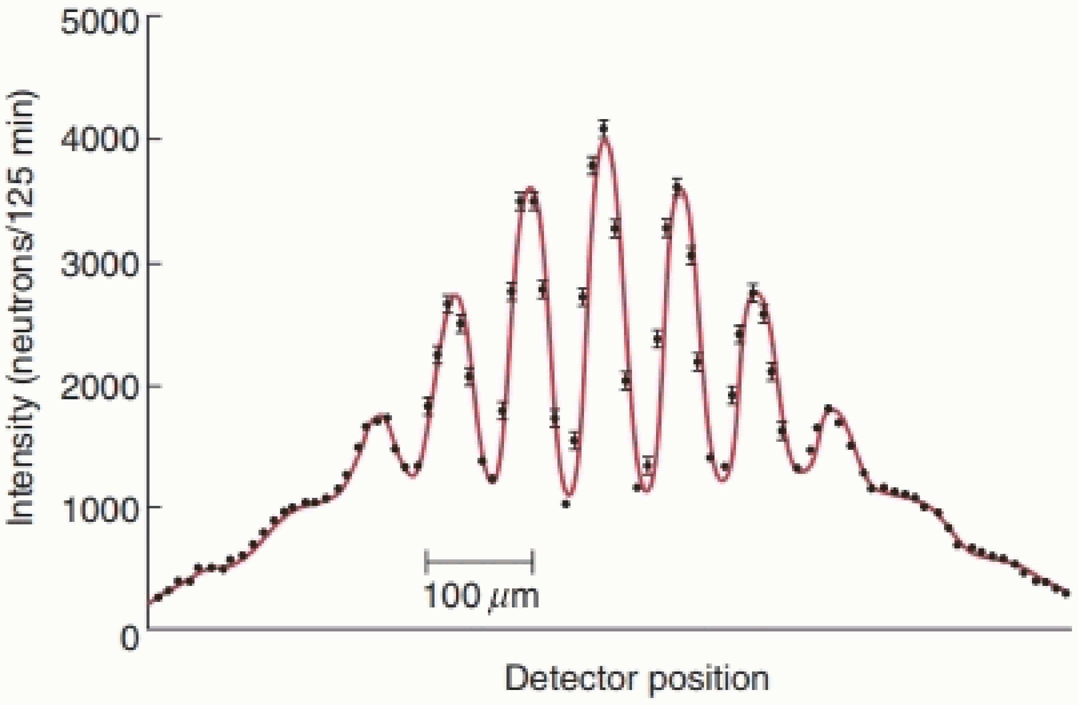


FIGURE 46-6. The interference pattern reported by Gähler and Zeilinger in 1991 when they allowed a beam of neutrons to fall on a screen containing a double slit. These results show that not only electrons but also neutrons are wave-like.

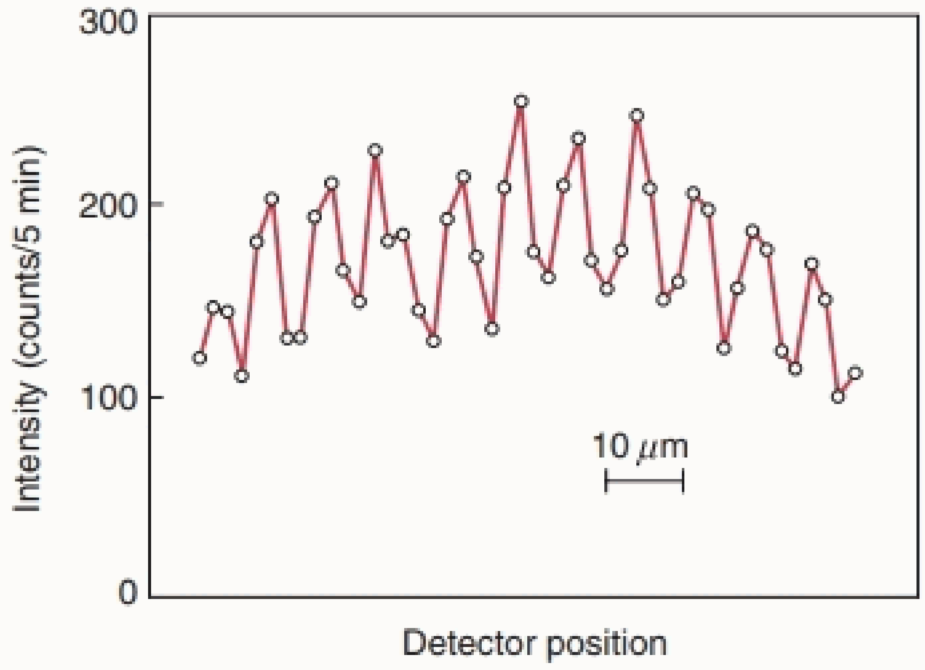


FIGURE 46-7. The interference pattern reported in 1991 by O. Carnal and J. Mlynek when they allowed a beam of helium atoms to fall on a screen containing a double slit. Not only electrons and neutrons but also helium atoms are wave-like.

hydrogen—that have few electrons. Neutrons, on the other hand, interact largely with the nucleus of the atom and can be used to fill this gap. Figure 46-8, for example, shows the structure of solid benzene (C_6H_6) as deduced from neutron diffraction studies. The six carbon atoms that form the familiar benzene ring are clearly shown, as are the six appended hydrogen atoms.

With high-energy protons or electrons, whose de Broglie wavelength is of the order of a nuclear diameter (10^{-14} m), it is even possible to observe the diffraction pattern of an atomic nucleus. From the diffraction pattern, the size and shape of the nucleus can be determined.

Taken together, these results for neutrons, protons, helium atoms, and other particles provide convincing evidence that the wave nature is a general property of particles and not only a property of electrons.

The Electron Microscope

Among the earliest practical applications of the wave nature of matter was the development of the electron microscope, which occurred within a decade of de Broglie's hypothesis. The resolving power of conventional optical microscopes is limited by the wavelength of the light employed; the smaller the wavelength, the finer is the detail that can be examined. Point objects separated by as little as about 200 nm can be resolved optically, using light of the shortest practical wavelength. However, as Sample Problem 46-1 shows, electron beams with wavelengths at least 100 times smaller than this can be produced.

A microscope based on electron beams became possible when it was discovered that a short compact solenoid, sheathed in iron, could bring to a focus an electron beam sent along its central axis, in much the same way as a glass lens can bring a light beam to a focus. The first electron microscope, employing two such *magnetic lenses,* was constructed in 1931 by the German physicist Ernst Ruska. Some 55 years after the event, Ruska (then at age 80) shared the 1986 Nobel Prize for this discovery. Today electron microscopes are in use in scientific laboratories of all kinds all over the world. Figure 46-9 is an image of a fruit fly, taken with a specialized instrument known as a *scanning electron microscope.*

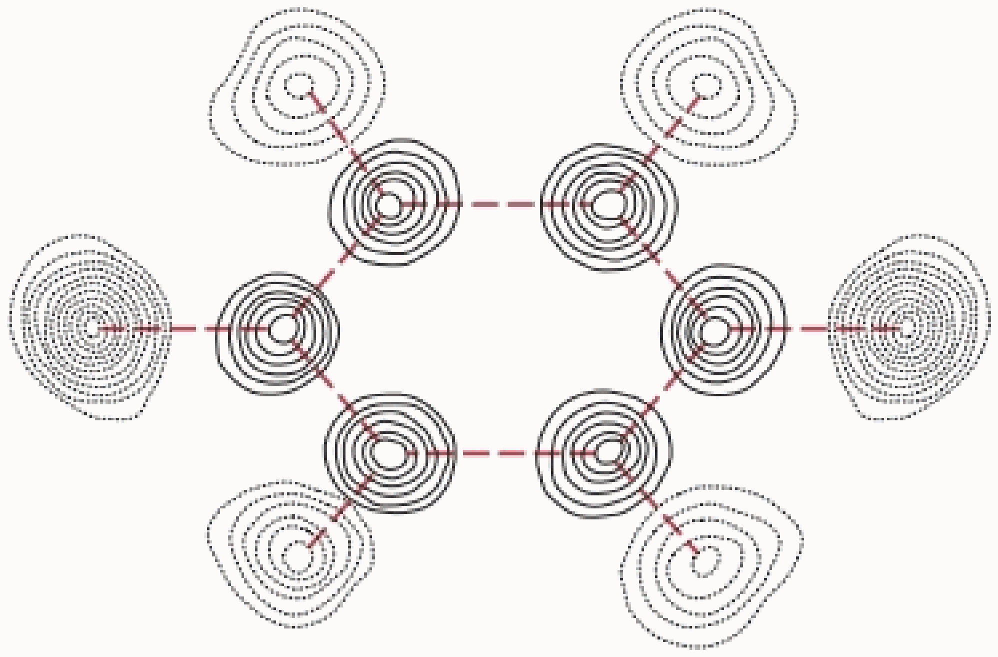


FIGURE 46-8. The atomic structure of solid benzene (C_6H_6) as deduced from neutron diffraction studies. The solid circles show the locations of the six carbon atoms that form the familiar benzene ring. The dotted circles show the six attached hydrogen atoms.

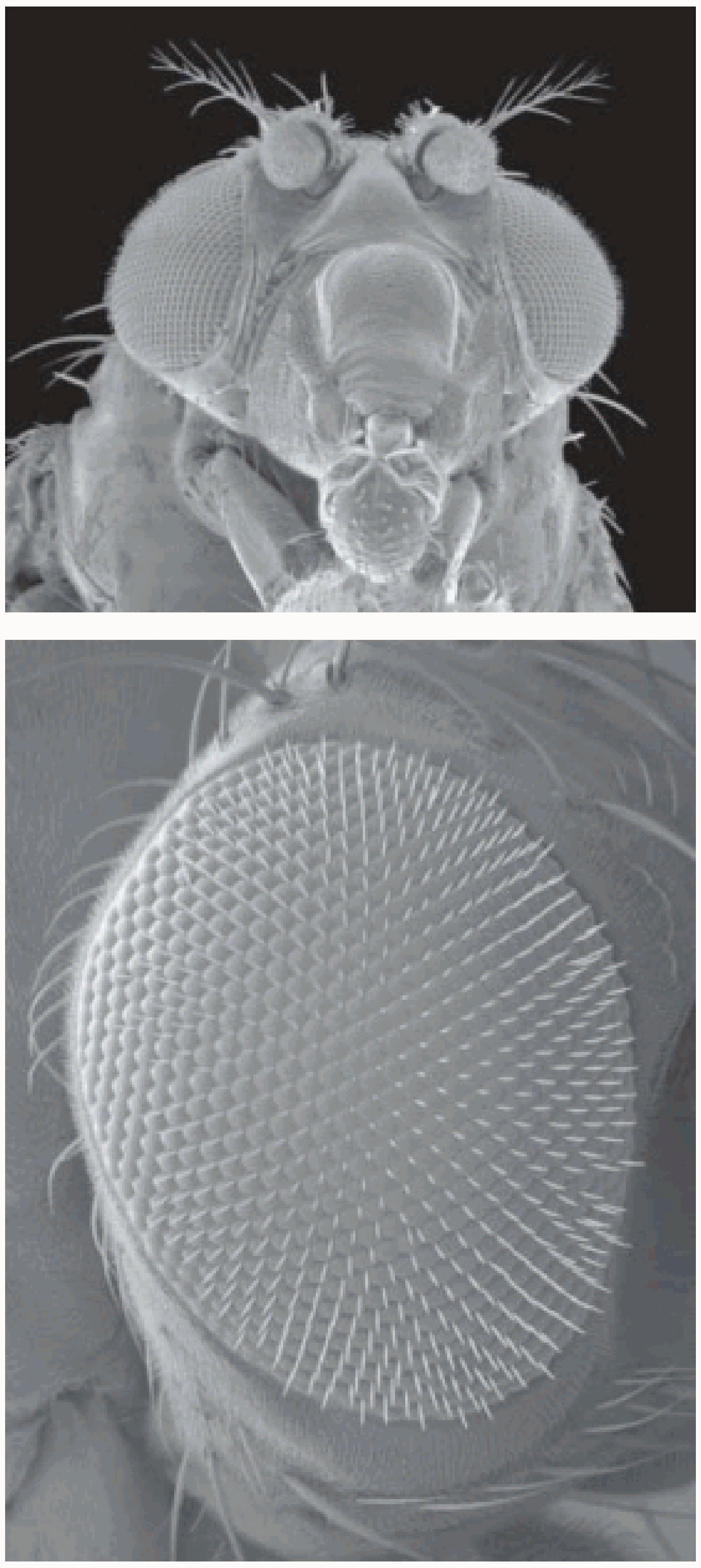


FIGURE 46-9. (Top) A fruit fly, photographed using a scanning electron microscope. (Bottom) The fly's eye.

It is important to realize that, in the electron microscope, the electrons in the beam can be analyzed as particles when their passage through the lens system is concerned. Their wave nature comes into play only in establishing the resolution that the microscope can achieve.

The Wave Nature of Single Electrons

When Tonomura and his associates reduced the intensity of the incident electron beam so that only one electron at a time passed through the apparatus of Fig. 46-4, the interference fringes—after a sufficiently long exposure—*still* appeared. In Section 45-1 we showed that photons demonstrated exactly this same behavior. Thus for electrons as well as for photons, experiment forces us to accept that each has both a particle and a wave nature. If we think of an electron as a particle, then we want to try to follow its trajectory and determine through which slit it passed; clearly a classical particle *cannot* pass through both slits. Yet for an interference pattern to be observed, we know that the electron (whatever it is!) *must* be passing through *both* slits. This is a behavior that we are perfectly happy to associate with waves, but are very reluctant to associate with particles.

It is our daily experience with macroscopic objects that forms our intuitive view of the "proper" behavior of matter. No object in our daily experience behaves in one circumstance like a particle and in another like a wave. It would be very difficult to hit a wave-like baseball, and we would be very surprised to see water waves suddenly delivering all their energy at one location in a concentrated particle-like bundle. We have no common daily experiences with the behavior of photons or electrons that would reveal their dual natures, and so their quantum behavior seems counterintuitive to us and is often said (even by some physicists!) to be "weird." Because intuition is bound to fail us when we ponder events in the quantum realm, it is not useful to try to imagine how an electron can go through two slits at the same time. In the words of Nobel laureate Richard Feynman:

> Do not keep saying to yourself, if you can possibly avoid it, "but how can it be like that" because you will get down the drain into a blind alley from which nobody has yet escaped. Nobody knows how it can be like that.

Its counterintuitive nature, however, does not prevent quantum mechanics from being the most successful theory that has ever been advanced. Every single one of its predictions that has been tested in the laboratory has passed with flying colors.

SAMPLE PROBLEM 46-2. In the double-slit experiment used for Fig. 46-7 to investigate the wave nature of a beam of helium atoms, the slits were 1 μm wide, the slit separation d was 8 μm, and the detector was a distance $D = 64$ cm from the plane of the slits. Measure the average spacing Δy between the fringes shown in Fig. 46-7 and calculate the de Broglie wavelength of the helium atoms in the beam that fell on the slits.

Solution In Section 41-2 (see Eq. 41-4) we found the spacing between adjacent interference fringes on the image screen of a double-slit interference experiment to be

$$\Delta y = \frac{\lambda D}{d}.$$

From Fig. 46-7, we estimate Δy, the average spacing between adjacent fringes, to be about 8 μm. Thus, from the above equation,

$$\begin{aligned}\lambda = \frac{d\,\Delta y}{D} &= \frac{(8 \times 10^{-6}\ \text{m})(8 \times 10^{-6}\ \text{m})}{0.64\ \text{m}} \\ &= 1.0 \times 10^{-10}\ \text{m} = 0.10\ \text{nm}.\end{aligned}$$

The value calculated from Eq. 46-1 was 0.103 nm, in excellent agreement with the above experimental value.

SAMPLE PROBLEM 46-3. Nuclear reactors designed for neutron diffraction experiments are built with a graphite cylinder piercing the shield wall, as shown in Fig. 46-10. After many collisions with the carbon atoms of the graphite, energetic neutrons from the hot reactor core come to thermal equilibrium with the graphite, which is essentially at room temperature ($T \approx 293$ K). The neutrons emerging from the cylinder, called *thermal neutrons,* have wavelengths in a range that is particularly useful in neutron diffraction studies. The most probable value of the de Broglie wavelength in such a distribution can be shown (see Problem 2) to be given by

$$\lambda_{\text{p}} = \frac{h}{\sqrt{5mkT}}, \tag{46-3}$$

in which m is the mass of the neutron, k is the Boltzmann constant, and T is the temperature in kelvins. (*a*) Find the most probable de Broglie wavelength in a beam of thermal neutrons. (*b*) A beam of these neutrons falls on a crystal C, in which the spacing between a certain set of atomic planes is $d = 0.304$ nm. An intense first-order Bragg reflection is observed from this set of atomic planes when the Bragg-scattering angle θ is as shown in Fig. 46-10. Find the angle θ.

Solution (*a*) Using Eq. 46-3, we have

$$\begin{aligned}\lambda_{\text{p}} &= \frac{6.63 \times 10^{-34}\ \text{J}\cdot\text{s}}{\sqrt{5(1.67 \times 10^{-27}\ \text{kg})(1.38 \times 10^{-23}\ \text{J/K})(293\ \text{K})}} \\ &= 1.14 \times 10^{-10}\ \text{m} = 0.114\ \text{nm}.\end{aligned}$$

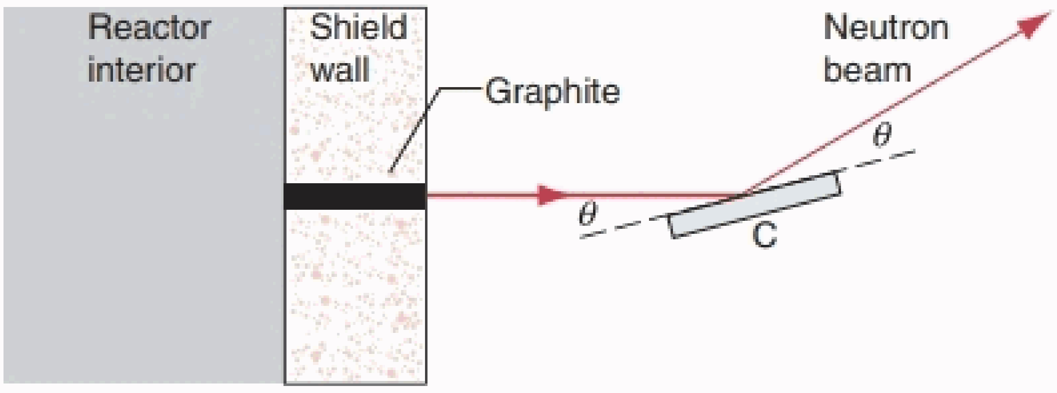


FIGURE 46-10. Sample Problem 46-3. An arrangement for generating a monoenergetic neutron beam for use in neutron diffraction studies. The neutron beam emerging from the reactor and falling on crystal C consists of thermal neutrons, with a substantial spread of energies. The beam reflected from the crystal is essentially monoenergetic.

(*b*) The Bragg formula for an x-ray reflection to occur from a set of atomic planes is Eq. 43-12,

$$2d \sin \theta = m\lambda \qquad (m = 1, 2, 3, \ldots),$$

in which d is the spacing between adjacent, parallel atomic planes and the integer m is the order of the reflected beam.

We can use this x-ray formula for particles such as neutrons if we interpret λ to be the de Broglie wavelength. Solving for the angle θ, we obtain for the first-order reflection ($m = 1$) of the thermal neutron beam whose most probable de Broglie wavelength we found in (*a*) above

$$\theta = \sin^{-1}\frac{m\lambda_{\rm p}}{2d} = \sin^{-1}\frac{(1)(0.114\ \text{nm})}{(2)(0.304\ \text{nm})} = 10.8°.$$

The arrangement of Fig. 46-10 is used to generate a monoenergetic beam of neutrons that can then be used for neutron diffraction studies, as was done to obtain the pattern shown in Fig. 46-8.

46-3 WAVES AND PARTICLES

As we have seen, the evidence that matter is wave-like is very strong. However, the evidence that matter is particle-like is equally as strong. How can we mix the wave and particle aspects into a consistent description of matter? We have discussed several techniques for determining the wavelength of a matter wave, but what does its amplitude represent?

We will begin a more rigorous mathematical discussion of the amplitude of the matter wave in Section 46-5, but for now we will assume that it carries information about the location of the particle: the wave has large amplitude where the particle is likely to be found, and it has small amplitude where the particle is unlikely to be found. If the wave has a constant amplitude throughout a region of space, the particle is equally likely to be found anywhere in that region. If the wave has zero amplitude throughout a region, the particle is never found there.

One property that we would like for particles (even particles with a wave-like nature) to have is the ability to be *localized.* That is, we would like to know where the particle is located. We may not know exactly, but we can get a rough idea. For example, we may know that an electron is attached to an atom of diameter 0.1 nm, indicating that we have localized the electron in that region of space. Or a neutron may pass through a narrow slit, allowing us (at least for an instant) to localize its position to a region the size of the slit width.

Let us see what insights we can gain by trying to put together a collection of waves in such a way that we obtain something that resembles the description of a localized particle.

Localizing a Wave Packet in Space

Figure 46-11*a* is a "snapshot" of a wave moving along the x axis, taken at an arbitrary time, say $t = 0$. The wave extends from $x = -\infty$ to $x = +\infty$ and has a sharply defined wavelength λ_0. Here, and in most of what follows, it will prove more convenient to work, not with the wavelength, but with the *wave number* k $(=2\pi/\lambda)$. As Fig. 46-11*b* shows, the infinite wave of Fig. 46-11*a* has a single, well-defined wave number k_0 $(=2\pi/\lambda_0)$. However, there is nothing about this wave that suggests the localization in space that we associate with a particle. Such a wave has no beginning, no end, and no single distinctive marker. If this wave were describing a particle, we would say that the particle could be found anywhere from $x = -\infty$ to $x = +\infty$; it is completely *un*localized.

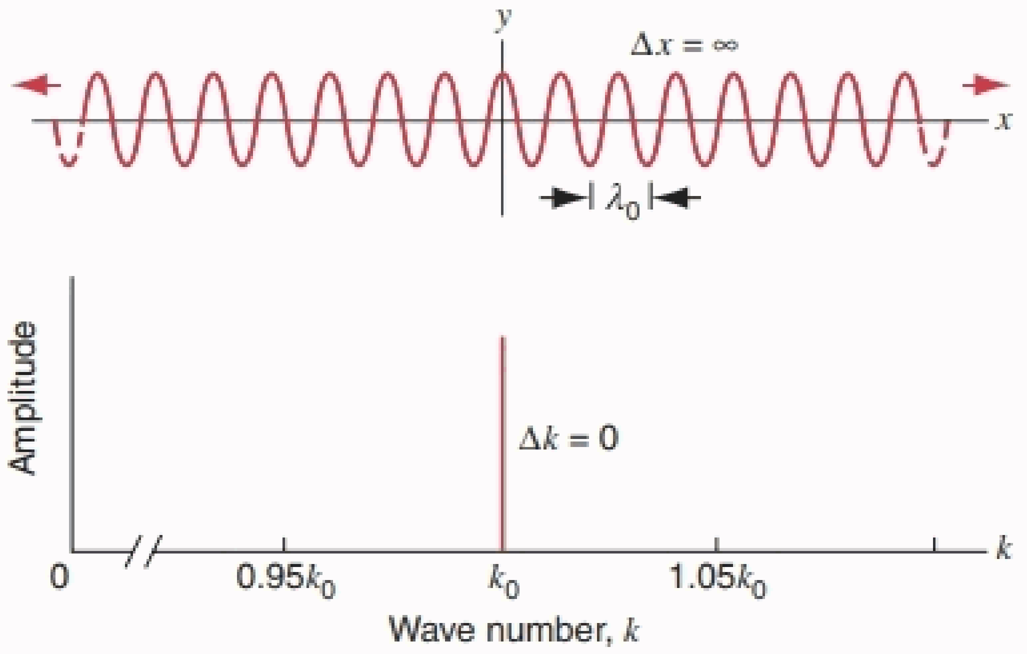


FIGURE 46-11. (*a*) A harmonic wave viewed at $t = 0$. (*b*) This (infinitely long) wave has a single wavelength λ_0 and, correspondingly, a single wave number k_0 $(=2\pi/\lambda_0)$.

As we learned in Section 18-7, we can create almost any wave shape we want by adding up sine or cosine waves with properly chosen wave numbers, amplitudes, and phases. Figure 46-12*a* shows a *wave packet* that could be built up in this way. This collection of many (infinitely

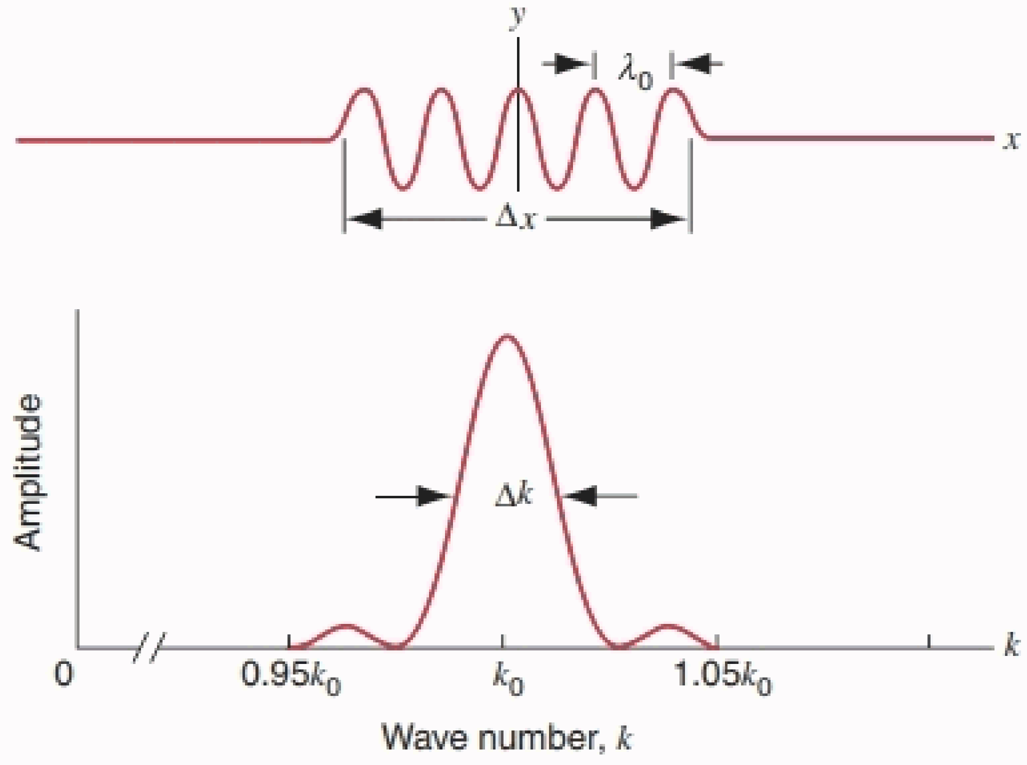


FIGURE 46-12. (*a*) A segment of a harmonic wave of length Δx—that is, a wave packet. The wave packet cannot be said to be sinusoidal because, unlike the wave of Fig. 46-11*a*, its wave-like feature does not extend to infinity. (*b*) The distribution of the wave numbers of the various harmonic components that combine to form the wave packet. The central peak has a width, measured at half its amplitude, of Δk. The maximum of the distribution occurs at $k = k_0 = 2\pi/\lambda_0$.

long) waves adds up to make a wave packet of length Δx and adds to zero everywhere else. We now have something that suggests a particle—namely, a localization in space measured by Δx. If this wave represents a particle, then the particle is likely to be found in the region of size Δx and unlikely to be found outside that region.

However, as Fig. 46-12*b* shows, our wave packet no longer contains a single wave number k_0 but rather a spread of wave numbers centered about k_0. Let Δk in Fig. 46-12*b* be a rough measure of the spread of wave numbers needed to form the packet of Fig. 46-12*a*. The (dimensionless) product of Δk and Δx proves to be of the order of unity, or

$$\Delta k \cdot \Delta x \cong 1. \tag{46-4}$$

This equation tells us that the sharper we wish the wave packet to be (that is, the smaller the value of Δx), the larger must be the range of wave numbers required to build it up. Conversely, the narrower the spread in Δk (or equivalently, the smaller the distribution of wavelengths), the less localized the particle will be.

Localizing a Wave Packet in Time

A true particle is localized in time as well as in space. We can visualize a wave packet built up in time if, in Fig. 46-12*a*, we replace x with t and the wave number k with the angular frequency ω $(=2\pi f)$. By analogy with Eq. 46-4, the duration Δt of our new wave packet is related to the spread $\Delta\omega$ of angular frequencies needed to make it up by

$$\Delta\omega \cdot \Delta t \cong 1. \tag{46-5}$$

Much of the information involved in telephone, television, radar, or computer links is sent from point to point in the form of wave packets. The electronic data handling pathways over which these pulses are transmitted should be sensitive over the full range of frequencies included in the pulses. Equation 46-5 tells us that the shorter the pulse, the greater must be the acceptable frequency range (that is, the *bandwidth*) of the system.

SAMPLE PROBLEM 46-4. A radar transmitter emits pulses of electromagnetic radiation 0.15 μs long at a wavelength of $\lambda = 1.2$ cm. (*a*) To what central frequency should the radar received be tuned? (*b*) What is the length of the wave packet? (*c*) What should be the bandwidth of the receiver?

Solution (*a*) The central frequency f_0 is given by

$$f_0 = \frac{c}{\lambda_0} = \frac{3.00 \times 10^8 \text{ m/s}}{0.012 \text{ m}} = 2.5 \times 10^{10} \text{ Hz} = 25 \text{ GHz}.$$

The central angular frequency is

$$\omega_0 = 2\pi f_0 = 2\pi(2.5 \times 10^{10} \text{ Hz}) = 1.6 \times 10^{11} \text{ rad/s}.$$

(*b*) The length of the wave packet (pulse) is

$$\Delta x = c\,\Delta t = (3.00 \times 10^8 \text{ m/s})(0.15 \times 10^{-6} \text{ s}) = 45 \text{ m}.$$

(*c*) The required bandwidth of the receiver is given approximately by Eq. 46-5, or

$$\Delta f = \frac{\Delta\omega}{2\pi} = \frac{1}{2\pi\,\Delta t} = \frac{1}{(2\pi)(0.15 \times 10^{-6} \text{ s})}$$
$$= 1.1 \times 10^6 \text{ Hz} = 1.1 \text{ MHz}.$$

46-4 HEISENBERG'S UNCERTAINTY PRINCIPLE

Using Eq. 46-1 ($\lambda = h/p$), we can write the wave number associated with a matter wave as

$$k = \frac{2\pi}{\lambda} = \frac{2\pi p}{h}, \tag{46-6}$$

in which p is the momentum of the associated particle. Consider a wave packet of matter waves of width Δx. According to the analysis illustrated in Fig. 46-12, such a pulse—of finite length—must be built up of an assembly of sine or cosine waves whose spread in wave number is Δk. From Eq. 46-6 we see that a spread in wave number implies a spread in momentum, or

$$\Delta k = \frac{2\pi}{h}\Delta p.$$

From Eq. 46-4 we then find

$$\Delta x \cdot \Delta k = \Delta x \cdot (2\pi/h)\,\Delta p \cong 1$$

or

$$\Delta x \cdot \Delta p \cong \frac{h}{2\pi}. \tag{46-7}$$

Taking into account that momentum is a vector, we can generalize this relationship to

$$\begin{aligned} \Delta x \cdot \Delta p_x &\geq h/2\pi, \\ \Delta y \cdot \Delta p_y &\geq h/2\pi, \\ \Delta z \cdot \Delta p_z &\geq h/2\pi. \end{aligned} \tag{46-8}$$

These are *Heisenberg's uncertainty relationships,* first derived by the German physicist Werner Heisenberg (1901–1976), one of the founders of quantum mechanics. They are the mathematical formulations of *Heisenberg's uncertainty principle,* which states:

> *It is not possible to measure, simultaneously, the position and the momentum of a particle with unlimited precision.*

Equations 46-8 tell us that the more precisely we try to pin down the position of a particle—that is, the smaller we seek to make Δx—the less we can learn about its momentum—that is, the larger will be Δp. The converse is also true. In particular, if we were able to assign an *absolutely precise* value to the particle's momentum ($\Delta p = 0$)

then we would not be able to assign a position at all; the particle could be found anywhere ($\Delta x = \infty$).

In Eq. 46-8 we have substituted an "equal to or greater than" symbol ($\geq$) for the "is approximately equal to" symbol of Eq. 46-7 ($\cong$). We did this to recognize that our measuring instruments, however refined, can never be ideal. The measured product $\Delta x \cdot \Delta p_x$ in Eq. 46-8, for example, will, in the real world, always be greater than the limits imposed by the uncertainty relationships. It can *never* be less.*

The message of the uncertainty principle is that there is a limit to the extent to which the concept of "particle" can be extended from the Newtonian world to the quantum world. In the quantum world, it is *wrong* to imagine that a particle really has a definite position and momentum but, for some reason, we are not able to measure them. It is *wrong* to visualize a particle as a tiny mass point moving along a path, with its position and velocity well defined at every instant. The very notion of "trajectory" belongs to the Newtonian world, not the quantum world. What, for example, can be the trajectory of a single photon or electron that encounters a double-slit experiment?

At the same time, well-defined tracks, created as energetic single particles pass through a liquid hydrogen bubble chamber (see Fig. 46-13), strongly suggest a trajectory. How to reconcile this evidence of our eyes with the wave concept? Figure 46-14, which suggests a particle detected at point A and detected again at point B, guides us to the answer. According to quantum mechanics, the wave describing the particle travels—with equal probability—on *all paths* connecting A and B, only a few of which are shown in the figure. In the double-slit experiment there were only two paths connecting A and B, all others being intercepted by the screen containing the slits. In the situation of Fig. 46-14, however, there is an infinite number of possible paths. It turns out that only for waves following paths that lie close to the straight line connecting A and B will constructive interference occur at point B. The wave describing the particle will have large amplitude only close to that line; that is where the particle is likely to be found, and that is what we observe as the trajectory of the particle. Waves following paths far from this line will interfere destructively; they will cancel. Thus, from the wave point of view, the trajectory results from the constructive and destructive interference of many waves that represent all possible paths.

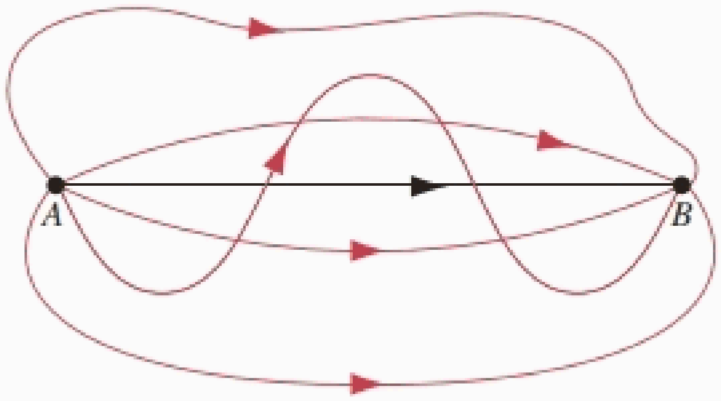


FIGURE 46-14. An electron is detected at A and then again at B. According to quantum mechanics, the matter waves representing the electron *follow all paths* connecting these two points. Only paths close to the straight-line path interfere constructively.

The Energy–Time Uncertainty Relationship

Thus far we have considered only the *wavelengths* of matter waves and have said nothing about their *frequencies.*

By analogy with Einstein's photon equation ($E = hf$), the uncertainty in the frequency of a matter wave is related to the uncertainty in the energy E of the corresponding particle by $\Delta f = \Delta E/h$. Substituting this into Eq. 46-5 with $\Delta\omega = 2\pi\,\Delta f$, we obtain

$$\Delta E \cdot \Delta t \geq h/2\pi, \tag{46-9}$$

which is the mathematical relationship of the uncertainty principle expressed in terms of different parameters. In words, it says:

> *It is not possible to determine both the energy and the time coordinate of a particle with unlimited precision.*

All energy measurements carry an inherent uncertainty unless you have an infinite time available for the measurement. In an atom, for example, the lowest energy state (the

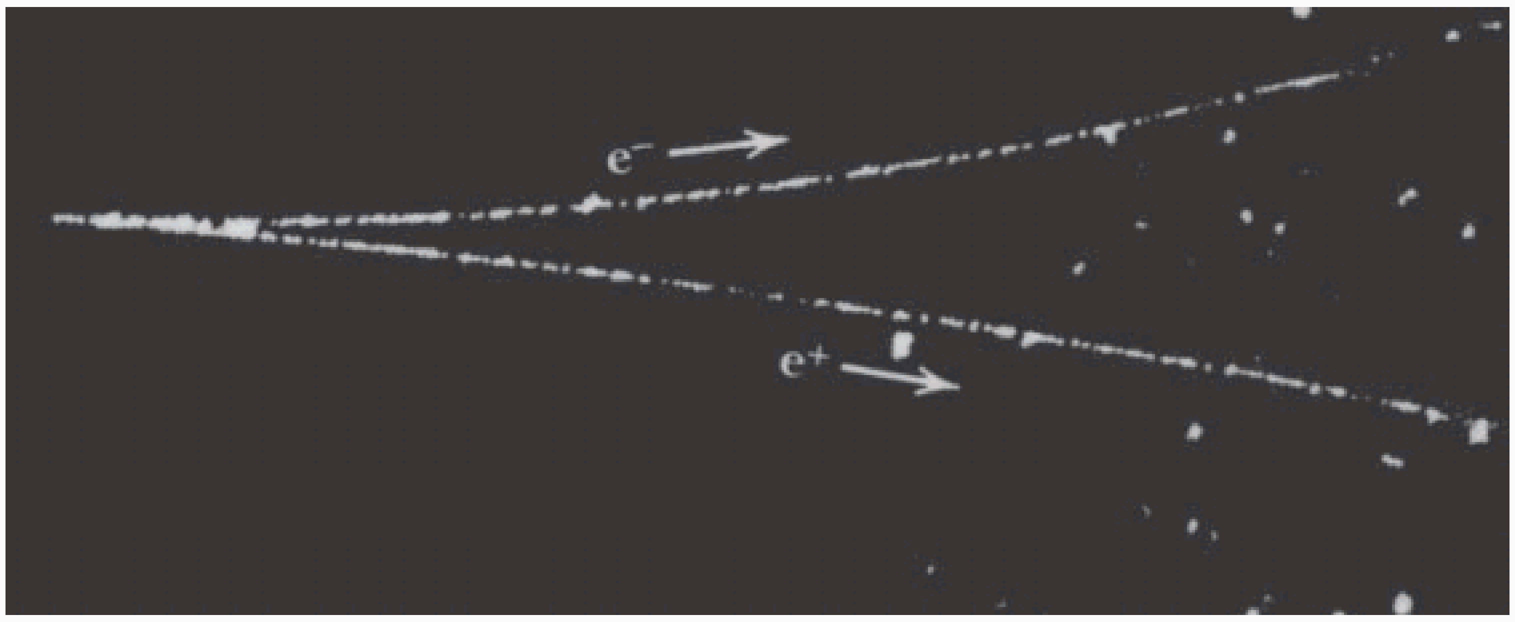


FIGURE 46-13. An electron and a positron, formed at the left vertex, pass through a bubble chamber filled with liquid hydrogen. Trails of bubbles mark their paths. The tracks are not straight because the particles are deflected by a magnetic field that permeates the chamber. How can such features be explained in terms of matter waves?

---

*A more rigorous derivation gives $h/4\pi$ in place of $h/2\pi$ in Eq. 46-8. However, the difference has no practical consequences for our discussions.

so-called ground state) has a well-defined energy because the atom normally exists indefinitely in that state. The energies of all states of higher energy (the excited states) are less precisely defined because the atom—sooner or later—will move spontaneously to a state of lower energy. On average, only a certain time Δt is available so that an energy measurement will be uncertain by an amount ΔE given by $(h/2\pi)/\Delta t$.

SAMPLE PROBLEM 46-5. (*a*) A 10-eV electron moves in the direction of increasing x with a speed of 1.88×10^6 m/s. Assume that you can measure this speed to a precision of 1.0%. With what precision can you simultaneously measure its momentum? (*b*) A golf ball has a mass of 45 g and a speed of 40 m/s, which you can measure with a precision of 1.0%. What limits does the uncertainty principle place on your ability to measure its position?

Solution (*a*) The electron's momentum is

$$p_x = mv_x = (9.11 \times 10^{-31}\ \text{kg})(1.88 \times 10^6\ \text{m/s})$$
$$= 1.71 \times 10^{-24}\ \text{kg}\cdot\text{m/s}.$$

The uncertainty Δp_x in momentum is 1.0% of this, or 1.71×10^{-26} kg · m/s. The uncertainty in position is then, from Eq. 46-8,

$$\Delta x \cong \frac{h}{(2\pi)(\Delta p_x)} = \frac{6.63 \times 10^{-34}\ \text{J}\cdot\text{s}}{(2\pi)(1.71 \times 10^{-26}\ \text{kg}\cdot\text{m/s})} = 6.2\ \text{nm}.$$

The size of a typical atom is about 0.1 nm, so Δx is about the length of a chain of 60 atoms. Given your measurement of the electron's momentum, there is simply no way you can simultaneously assign a position to the electron to a better precision than this.

(*b*) This example is exactly like part (*a*) except that the momentum of the golf ball is very much larger than that of the electron. The same calculation yields, for the golf ball, $\Delta x \cong 6 \times 10^{-33}$ m. This is a very small distance indeed, about 10^{18} times smaller than a typical atomic nucleus. Where macroscopic objects are concerned, the uncertainty principle sets no meaningful limit to the precision of measurement. You could never have discovered that principle by careful measurement of flying golf balls or speeding bullets!

46-5 THE WAVE FUNCTION

Equation 46-1 shows us how to find the wavelength of the matter wave associated with a particle if we know its momentum. We now ask: How do we describe the amplitude of a matter wave? Put less formally: Just what is it that is waving?

The displacement function for a transverse wave traveling along a stretched string in the direction of increasing x (see Section 18-3) is given by

$$y(x, t) = y_0 \sin(kx - \omega t). \tag{46-10}$$

Here $k\ (= 2\pi/\lambda)$ is the wave number, $\omega\ (= 2\pi f)$ is the angular frequency, and y_0 is the wave amplitude. Here and in most of what follows, it will be more convenient to deal with k (rather than λ) and with ω (rather than f).

Consider now the matter wave associated with a particle of mass m traveling in the direction of increasing x and on which no force acts, a so-called *free particle.* To describe the displacement associated with such a wave, the Austrian physicist Erwin Schrödinger (1887–1961) introduced a quantity $\Psi(x, t)$ that he called the *wave function.*

In Section 46-6 we will show that the wave function for a free particle moving in the direction of increasing x is given by

$$\Psi(x, t) = \psi_0\, e^{i(kx - \omega t)}, \tag{46-11}$$

in which ψ_0 is the amplitude of the wave. Note that the wave function contains the *imaginary number* $i\ (= \sqrt{-1})$. $\Psi(x, t)$ is a *complex quantity.*

At first glance, the form of this expression for a traveling matter wave seems quite different than that given in Eq. 46-10 for a traveling wave on a string. Recall, however, that *any* function of $(kx - \omega t)$ can describe a wave traveling in the direction of increasing x. Also, the mathematical identity

$$e^{i\theta} \equiv \cos\theta + i\sin\theta, \tag{46-12}$$

in which θ is any angle, shows that the exponential form of Eq. 46-11 is less different than you might think from the trigonometric form of Eq. 46-10.

The physical interpretation of the wave function was given by the German physicist Max Born (1882–1970). He asserted that physical meaning should be given not to Ψ itself, but to the product of Ψ and its *complex conjugate* Ψ^*. (Recall that the complex conjugate of a complex number is that number with i replaced by $-i$, wherever it occurs.) Specifically, Born postulated

> *The product $\Psi\Psi^*\, dx$ gives the probability that the particle in question will be found between positions x and $x + dx$.*

In the quantum world we cannot say *where* a particle is; we can only say where it is *likely to be.* We call the product $\Psi\Psi^*$ the *probability density,* symbol $P(x)$, so that

$$P(x) = \Psi\Psi^*. \tag{46-13}$$

Even though the wave function $\Psi(x, t)$ is usually complex, the probability density will always be a real number (positive or zero) just as it must be if the probability density is to have physical meaning. Using Eq. 46-13, the probability density for the free particle described by Eq. 46-11 is

$$P(x) = [\psi_0\, e^{i(kx-\omega t)}][\psi_0^*\, e^{-i(kx-\omega t)}] = |\psi_0|^2, \tag{46-14}$$

which is a constant, independent of x or t. Figure 46-15 shows a plot of $P(x)$ at any specified instant—say, $t = 0$. Because $P(x)$ is a constant for a free particle, we conclude that the particle can be found with equal probability at any point along the x direction, from $x = -\infty$ to $x = +\infty$.

This inability to pin down the location of a free particle is in complete accord with Heisenberg's uncertainty princi-

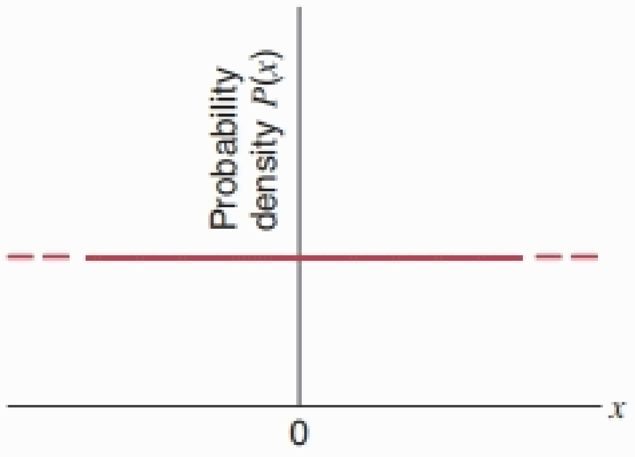


FIGURE 46-15. The probability density $\Psi\Psi^*$ for a free particle moving in the direction of increasing x. The particle has the same probability of detection at all points along the x axis.

ple. Because no force acts on a free particle, its momentum must be a precisely determined constant, so that $\Delta p_x = 0$. Heisenberg's uncertainty principle (Eq. 46-8) then predicts that $\Delta x = \infty$, which is exactly what Fig. 46-15 shows.

46-6 SCHRÖDINGER'S EQUATION

In Section 46-5, we stated without proof that Eq. 46-11 is the wave function that describes a free particle. How do we know that? The answer is that this equation is a solution of *Schrödinger's equation,* which is the fundamental equation of quantum mechanics advanced in 1926 by Erwin Schrödinger, who shared the 1933 Nobel Prize for this achievement. (There are other equivalent ways of formulating quantum mechanics; we have chosen to follow in Schrödinger's footsteps.)

Before we display Schrödinger's equation let us note that we can rewrite Eq. 46-11 as

$$\Psi(x, t) = (\psi_0 e^{+ikx})(e^{-i\omega t}),$$

in which, as we see, the wave function $\Psi(x, t)$ appears as the product of two factors, one involving only x and the other only t.

To sum up, we can write our wave function in the form

$$\Psi(x, t) = \psi(x)\, e^{-i\omega t}. \qquad (46\text{-}15)$$

In the rest of this chapter, we will focus our attention entirely on $\psi(x)$, the space-dependent portion of the wave function. Keep in mind that upper-case Greek letter psi (Ψ) involves both position and time and lower-case psi (ψ) involves the position coordinate only.* Comparing Eq. 46-11 with Eq. 46-15 shows that, for a free particle,

$$\psi(x) = \psi_0\, e^{+ikx}. \qquad (46\text{-}16)$$

You can easily show that $\psi\psi^* = |\psi_0|^2 = P(x)$, so that $\psi\psi^*$ also (see Eq. 46-14) gives the probability density of a traveling matter wave.

Now for Schrödinger's equation. For a particle (not necessarily free) traveling in the x direction, it is

$$-\frac{h^2}{8\pi^2 m}\frac{d^2\psi(x)}{dx^2} + U(x)\,\psi(x) = E\,\psi(x), \quad (46\text{-}17)$$

in which E is the (constant) total energy of the particle and $U(x)$ is its potential energy. We assume the usual situation—namely, that the potential energy does not depend on time.

We cannot derive Schrödinger's equation from more basic principles. It *is* the basic principle, in the same way that Newton's laws of motion are the basic principles for Newtonian mechanics and Maxwell's equations are the basic principles for electromagnetism. In all three cases we must accept the basic equations provisionally and see whether or not they lead to predictions that are confirmed by experiment. In all three cases, they do.

A Free Particle

If the particle is a free particle, its potential energy $U(x)$ is a constant which we can take to be zero for all values of x. The total energy E of the moving particle must then be entirely kinetic. That is, we must have $E = K = p^2/2m$ in which p is the particle's momentum. With this substitution, Schrödinger's equation (Eq. 46-17) then becomes

$$\frac{h^2}{4\pi^2 p^2}\frac{d^2\psi(x)}{dx^2} + \psi(x) = 0. \qquad (46\text{-}18)$$

Earlier, we stated without proof that the wave function $\psi(x)$ of Eq. 46-16 describes a free particle moving in the direction of increasing x. Let us now prove this claim by substituting Eq. 46-16 and its second derivative into Schrödinger's equation, Eq. 46-18.

The wave function we seek to test is

$$\psi(x) = \psi_0\, e^{ikx}.$$

Its first derivative is

$$\frac{d\psi(x)}{dx} = \psi_0\,(ik)\,e^{ikx}$$

and its second derivative (recalling that $i^2 = -1$) is

$$\frac{d^2\psi(x)}{dx^2} = \psi_0\,(-k^2)e^{ikx} = -k^2\,\psi(x).$$

If we substitute this expression for $d^2\psi(x)/dx^2$ into Eq. 46-18 we find

$$-\frac{h^2}{4\pi^2 p^2}k^2\,\psi(x) + \psi(x) = 0.$$

Canceling $\psi(x)$ and using Eq. 46-1 to substitute λ for p/h yields

$$k = \frac{2\pi}{\lambda},$$

---

*Formally, $\Psi(x, t)$ is called a *wave function* and $\psi(x)$ an *eigenfunction.* We will not use that latter notation here.

an identity. Because substituting Eq. 46-16 into Schrödinger's equation leads to an identity, we have proved that Eq. 46-16 is indeed a solution of that equation.

46-7 BARRIER TUNNELING

If you roll a marble across the floor toward a brick wall, you expect the marble to rebound from the wall. You do not expect it to appear on the other side of the wall. However, quantum mechanics predicts that behavior like this does happen for electrons and other particles with small masses. Let us see what Schrödinger's equation can tell us about this phenomenon.

Figure 46-16*a* suggests an electron of (constant) total energy E moving in the direction of increasing x. Its potential energy is zero except when the electron is in the region $0 < x < L$. In that region its potential energy is U_0, where $U_0 > E$. We define this region as a *potential energy barrier* (or, loosely, a *potential barrier*) of height U_0 and thickness L.

If the electron obeyed Newtonian mechanics, it would rebound from the barrier and would move back in the direction from which it came, just as a marble would rebound from a brick wall. However, electrons obey the rules of quantum mechanics and, according to those rules, the electron will indeed rebound from the barrier, but there is also a finite chance that it will also "leak through" the barrier by a process called *barrier tunneling* and appear on the other side.

Because the potential energy changes abruptly—at $x = 0$ and also at $x = L$—we must seek solutions of Schrödinger's equation (Eq. 46-16) separately in three regions: (1) to the left of the barrier, (2) within the barrier, and (3) to the right of the barrier. We must then choose the arbitrary constants that appear in these three solutions so that the values of $\psi(x)$ join smoothly (no jumps or kinks) at $x = 0$ and at $x = L$. Figure 46-16*b* shows a plot of the probability density $P(x)$ that we can compute after we have carried out all this solving and matching. Let us now examine the solutions of Schrödinger's equation in the three regions of interest.

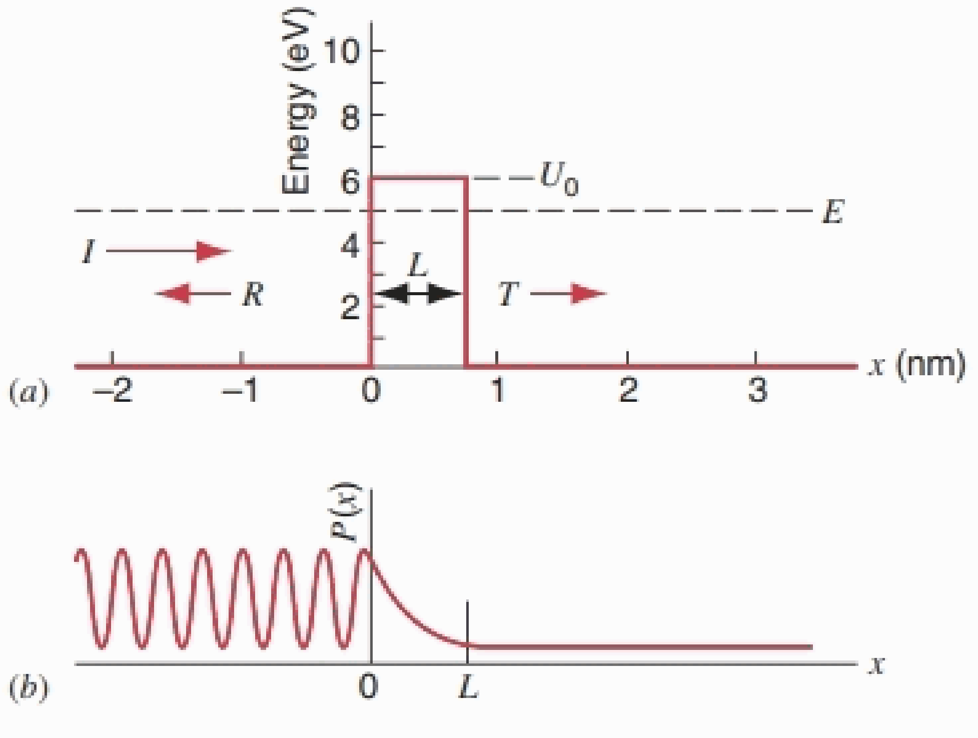


FIGURE 46-16. (*a*) An electron moves toward a potential barrier from the left, with constant total energy E. Within the barrier, its potential energy U_0 exceeds its total energy. (*b*) The probability density of the matter wave representing the electron. To the left of the barrier, the incident wave I and the wave R reflected (with lower intensity) from the barrier interfere with each other. The transmitted wave T appears to the right of the barrier. Within the barrier, the probability density decays exponentially.

To the left of the barrier: The solution of Schrödinger's equation is

$$\psi(x) = \psi_1 e^{ikx} + \psi_2 e^{-ikx} \qquad (x < 0), \qquad (46\text{-}19)$$

in which ψ_1 and ψ_2 are amplitude constants. The first term of Eq. 46-19 (compare Eq. 46-16) represents the incoming matter wave, which is traveling in the direction of increasing x. The second term represents the wave reflected from the barrier and thus traveling in the direction of decreasing x. As Sample Problem 46-7 shows, these two waves, which travel in opposite directions, interfere with each other and generate a standing wave pattern to the left of the barrier. The minima of this pattern are not zero because the reflected wave has a smaller amplitude than the incident wave. That is, because part of the incident wave is transmitted, $|\psi_2| < |\psi_1|$.

To the right of the barrier: On the far side of the barrier, the transmitted wave is given by Eq. 46-16, or

$$\psi(x) = \psi_5 e^{ikx} \qquad (x > L), \qquad (46\text{-}20)$$

in which ψ_5 is an amplitude constant. Compare this part of Fig. 46-16*b* with Fig. 46-15. In each case the probability density is constant; that is, the probability of finding the electron (however small that probability may be) is the same for all values of x in the region $x > L$.

Within the barrier: Here Schrödinger's equation (Eq. 46-17) is

$$-\frac{h^2}{8\pi^2 m}\frac{d^2\psi}{dx^2} + [U_0 - E]\,\psi = 0. \qquad (46\text{-}21)$$

Note that, because $U_0 > E$, the quantity in the square brackets above is positive.

Let us try a solution of the form

$$\psi(x) = \psi_3 e^{-k'x} + \psi_4 e^{+k'x} \qquad (46\text{-}22)$$

in which ψ_3 and ψ_4 are amplitude constants and, as you can see, the imaginary number i does not appear. You can easily show that if this function and its second derivative are substituted into Eq. 46-21 the variable x will be eliminated, provided that the constant k' in Eq. 46-22 has the value

$$k' = \frac{2\pi\sqrt{2m\,(U_0 - E)}}{h}. \qquad (46\text{-}23)$$

The detailed solution shows $|\psi_3| \gg |\psi_4|$, and the probability density that follows from Equation 46-22 is then

$$P(x) = \psi\psi^* \approx |\psi_3|^2 e^{-2k'x}. \qquad \text{(46-24)}$$

This quantity decreases with x. Thus, although the particle may be found in the classically forbidden region for which $U_0 > E$, you are less and less likely to find it the deeper you penetrate into that region. Finding a particle in a region where Newtonian physics forbids it to be is a common feature of quantum mechanics.

The Transmission Coefficient

We can define a transmission coefficient T for the particle and the barrier as the ratio of the intensity of the transmitted wave, $|\psi_5|^2$, to that of the incident wave, $|\psi_1|^2$. It can be shown that T is given by

$$T = 16\frac{E}{U_0}\left(1 - \frac{E}{U_0}\right)e^{-2k'L}, \qquad \text{(46-25)}$$

in which k' has the value given by Eq. 46-23. This formula is an approximation that holds only for barriers that are high and/or thick enough so that the transmission coefficient is small—that is, that $T \ll 1$.

The dominant factor in the transmission coefficient is the exponential, $e^{-2k'L}$. As a result, the value of T is very sensitive to the thickness L of the barrier and to the quantity k', which, in turn, depends on the mass m of the particle and the height U_0 of the barrier. The transmission coefficient decreases rapidly as the mass of the particle increases, becoming vanishingly small indeed as we progress from electrons to, say, marbles. If the barrier is infinitely tall, then $k' \to \infty$ and $T \to 0$, which suggests that the particle is 100% reflected from the barrier.

Some Examples

The barrier tunneling of matter waves has many practical applications. For a simple example, consider a bare copper wire that has been cut and the two ends twisted together. It still conducts electricity readily, even though the wires are unavoidably coated with a thin layer of copper oxide, an insulator. How do the electrons get through this (extremely thin) oxide layer? By barrier tunneling.

For a more exotic example, consider the core of the Sun, where the Sun's energy is being generated by thermonuclear fusion processes. Such processes involve the fusing together of light nuclei to form heavier ones, with the release of energy. Suppose that two protons are rushing together at high speed. They must get extremely close before their strong attractive nuclear forces can come into play and cause them to fuse. Meanwhile, they are slowed down by the repulsive Coulomb force that tends to drive them apart. They are, we can say, separated by a Coulomb barrier. The likelihood of fusion depends critically on the ability of the protons to tunnel through this barrier rather than surmount it. Without barrier tunneling the solar furnace would shut down.

Barrier tunneling plays a role, not only when two positively charged particles rush together but also when they fly apart, breaking the strong attractive nuclear bonds that hold them together and tunneling the Coulomb barrier in the opposite direction. The spontaneous emission of (positively charged) alpha particles by radioactive nuclei and the fission of heavy nuclei into two large fragments are two examples.

Among practical applications is the *tunnel diode,* in which the flow of electrons (by barrier tunneling) through a device can be rapidly turned on or off by controlling the height of the barrier, by varying an externally applied voltage, for instance. This can be done very rapidly, with response times of the order of 10^{-11} s or 10 ps possible. The 1973 Nobel Prize was shared by three "tunnelers," Leo Esaki (discoverer of the tunnel diode), Ivar Giaver (tunneling in superconductors), and Brian Josephson (discoverer of the Josephson junction.)

The Scanning Tunneling Microscope

In a *scanning tunneling microscope* (STM),* a specially prepared fine needle tip is scanned in a TV-like raster pattern over the surface of the sample being investigated; see Fig. 46-17. A potential difference of a few millivolts is applied between the needle tip and the sample, establishing a potential energy barrier. Electrons from the sample tunnel through the gap between the sample and the needle tip as a "tunnel current."

Normally, this tunnel current would vary widely as the gap between the needle and the surface structures changed during the scan. However, an electronic feedback mechanism is provided that automatically moves the needle up or down during the scan, so as to keep the tunnel

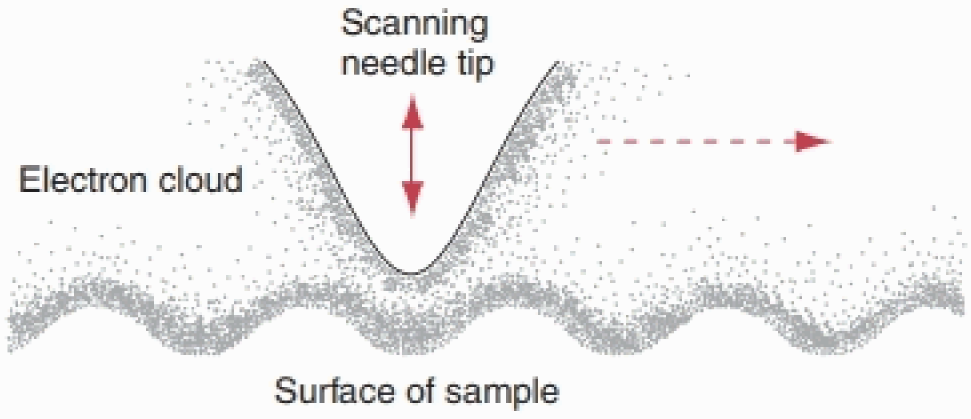


FIGURE 46-17. A sharp needle tip is scanned over the surface of a sample in a scanning tunneling microscope (STM).

---

*See "The Scanning Tunneling Microscope," by Gerd Binning and Heinrich Rohrer, *Scientific American,* August 1985. See http://www.almaden.ibm.com/vis/stm/lobby.html for a gallery of the STM and other photos.

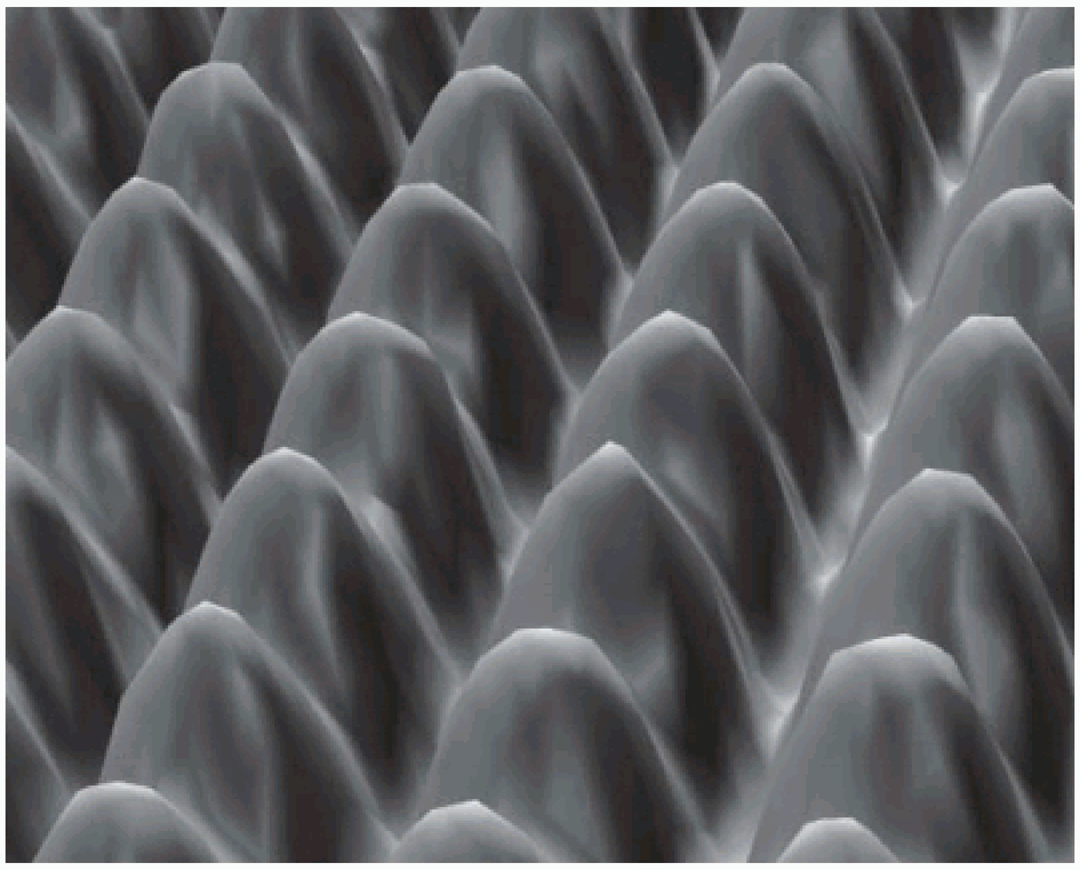


FIGURE 46-18. The regular arrangement of atoms that form the surface of a sample of nickel shows up clearly in this scanning tunneling microscope plot.

current—and thus the gap—constant. The needle tip's vertical position can then be displayed on a screen as a function of its location, producing a three-dimensional plot of the contour of the surface. Figure 46-18, which shows the regular arrangement of atoms forming the surface of a sample of nickel, is a typical example. In 1986, Gerd Binnig and Heinrich Rohrer of the IBM Zurich Research Laboratory shared the Nobel Prize (with Ernst Ruska, the inventor of the electron microscope) for the development of this device.

As Fig. 46-18 suggests, the motion of the needle tip, in both the horizontal plane and in the vertical direction, must be controllable with a reproducibility that is a small fraction of the dimensions of a single atom. This is accomplished by fastening the needle to the intersection of three ceramic rods, as Fig. 46-19 shows. The material of the rods has an interesting property (discovered by the brothers Pierre and Jacques Curie in 1880–81) called *piezoelectricity:* When a potential difference is applied across a rod, the dimensions of the rod change by a tiny but controllable amount, the change in length being directly proportional to the applied potential difference. By this technique the scanning motion can be controlled electronically and the vertical position of the tip can be coordinated with the position of the tip in the horizontal plane. Vertical displacements can be displayed on an expanded scale and color can be added to the display to bring out certain features.

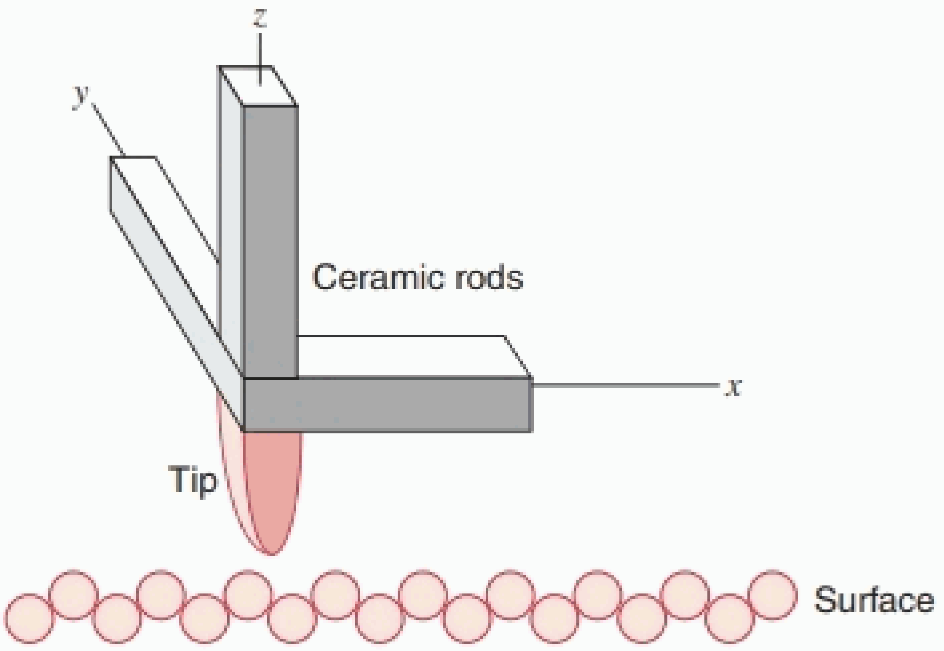


FIGURE 46-19. The needle tip in a scanning tunneling microscope is fastened to the intersection of three ceramic rods. By means of the piezoelectric effect, the tip can be scanned in the xy plane and also raised and lowered in the z direction.

SAMPLE PROBLEM 46-6. Figure 46-16*a* suggests an electron whose total energy E is 5.0 eV approaching a barrier whose height U_0 is 6.0 eV. The barrier thickness L is 0.70 nm. (*a*) What is the de Broglie wavelength of the incident electron? (*b*) What transmission coefficient T follows from Eq. 46-25? (*c*) What would be the transmission coefficient if the barrier thickness were reduced to 0.35 nm? If its height were increased to 7.0 eV? If the incident particle were a proton?

Solution (*a*) Before the electron reaches the barrier, its potential energy in the region through which it travels is zero. Proceeding as in Sample Problem 46-1 we find $\lambda = 0.55$ nm. Thus the barrier is 0.70/0.55 or about 1.3 de Broglie wavelengths thick.

(*b*) We have, from Eq. 46-23,

$$k' = \frac{2\pi\sqrt{2m(U_0 - E)}}{h}$$

$$= \frac{(2\pi)\sqrt{(2)(9.11\times10^{-31}\text{ kg})(6.0\text{ eV} - 5.0\text{ eV})(1.60\times10^{-19}\text{ J/eV})}}{6.63\times10^{-34}\text{ J}\cdot\text{s}}$$

$$= 5.12\times10^{9}\text{ m}^{-1}.$$

The quantity $k'L$ is then $(5.12\times10^{9}\text{ m}^{-1})(0.70\times10^{-9}\text{ m}) = 3.58$. Using Eq. 46-25, the transmission coefficient proves to be

$$T = 16\frac{E}{U_0}\left(1 - \frac{E}{U_0}\right)e^{-2k'L}$$

$$= (16)\frac{5.0\text{ eV}}{6.0\text{ eV}}\left(1 - \frac{5.0\text{ eV}}{6.0\text{ eV}}\right)e^{-2(3.58)} = 1.73\times10^{-3}$$

Of every 100,000 electrons that strike the barrier, only 173 will tunnel through it.

(*c*) Making appropriate changes (one at a time) in the solution to part (*b*) we find:

$L = 0.35$ nm yields $T = 0.062$
$U_0 = 7.0$ eV yields $T = 1.3\times10^{-4}$
$m = 1836m_e$ yields $T = 8.8\times10^{-134}$

Compare these newly calculated quantities to the value of $T\,(= 1.73\times10^{-3})$ calculated in part (*b*). Note that it is easier for the electron to penetrate the thinner barrier but more difficult for it to penetrate the higher barrier. The more massive proton penetrates hardly at all. Imagine how small T would be for a marble!

SAMPLE PROBLEM 46-7. Equation 46-19 gives the space-varying portion of the wave function that describes the matter wave to the left of the barrier of Fig. 46-16*a*—that is, in the region $x < 0$. Calculate the corresponding probability density in this region.

Solution Using Eq. 46-19 for $\psi(x)$, the probability density becomes

$$\begin{aligned} P(x) &= \psi\psi^* \\ &= (\psi_1\, e^{+ikx} + \psi_2\, e^{-ikx})(\psi_1^*\, e^{-ikx} + \psi_2^*\, e^{+ikx}) \\ &= |\psi_1|^2 + |\psi_2|^2 + \psi_1\psi_2^*\, e^{+2ikx} + \psi_1^*\psi_2\, e^{-2ikx}. \end{aligned}$$

Recall that, to form ψ^*, we have replaced i in ψ by $-i$, wherever it occurs.

The amplitudes ψ_1 and ψ_2 refer, respectively, to the incoming and the reflected components of the matter wave. Because the reflection is not complete (the wave is partially transmitted), the reflected wave must have a smaller amplitude than the incoming wave. Let us define a dimensionless *reflection parameter* $\rho = |\psi_2|/|\psi_1|$ in which $0 < \rho < 1$. After quite a bit of algebraic manipulation, we obtain

$$P(x) = |\psi_1|^2\,[1 + \rho^2 + 2\rho\cos(2kx + \Delta\psi)],$$

where $\Delta\psi$ is an unimportant phase angle that emerges from the calculation. Note that $P(x)$ is a real number, as all probability densities must be. It varies between a maximum value [for $\cos(2kx + \Delta\psi) = +1$] of

$$P_{\max} = |\psi_1|^2(1 + \rho)^2$$

and a minimum value [for $\cos(2kx + \Delta\psi) = -1$] of

$$P_{\min} = |\psi_1|^2(1 - \rho)^2.$$

Our results are in full agreement with Fig. 46-16*b*.

The distance Δx between adjacent maxima (or minima) occurs when the quantity $2kx$ in the cosine changes by 2π, so that $2k\,\Delta x = 2\pi$, or

$$\Delta x = \frac{\pi}{k} = \frac{\pi}{(2\pi/\lambda)} = \frac{\lambda}{2}.$$

Thus adjacent maxima (or minima) of the probability density plot of Fig. 46-16*b* are separated by one-half the de Broglie wavelength of the incident wave.

MULTIPLE CHOICE

46-1 Matter Waves

1. If the particles listed all have the same kinetic energy, which has the shortest wavelength?

(A) Electron (B) α particle
(C) Neutron (D) Proton

2. If the particles listed all have the same de Broglie wavelength, which has the largest kinetic energy?

(A) Electron (B) α particle
(C) Neutron (D) Proton

3. How does the de Broglie wavelength λ associated with a particle compare to the size of the particle?

(A) λ must be larger than the particle.
(B) λ must be smaller than the particle.
(C) λ could be larger or smaller than the particle.

46-2 Testing de Broglie's Hypothesis

4. A beam of electrons with speed v_0 passes through double slits and then is allowed to strike a fluorescent screen. An interference pattern is observed on the screen.

(*a*) The speed of the electrons is increased to $2v_0$. What happens to the spacing of the interference fringes on the screen?

(A) The spacing increases.
(B) The spacing decreases.
(C) The spacing remains the same.

(*b*) The electrons are instead replaced by protons moving with speed v_0. Compared to electrons at the same speed, what happens to the spacing of the interference fringes on the screen?

(A) The spacing increases.
(B) The spacing decreases.
(C) The spacing remains the same.

5. A single electron with de Broglie wavelength λ passes through a slit of width $d = 2\lambda$. The electron can then strike a fluorescent screen. What will be observed on the screen?

(A) A diffraction pattern
(B) A single flash as if the electron had moved in a straight line through the slit.
(C) A single flash that could occur anywhere on the screen.
(D) A single flash that would most likely occur where a corresponding diffraction pattern would have the highest intensity.

46-3 Waves and Particles

46-4 Heisenberg's Uncertainty Principle

6. An electron and a proton are both confined in boxes of length a. Measurements are made on the momentum of both particles. Which statement is true about the precision of these measurements?

(A) The electron's momentum can be measured more precisely than the proton's momentum.
(B) The proton's momentum can be measured more precisely than the electron's momentum.
(C) The electron's momentum can be measured only as precisely as the proton's momentum.

7. The position of an electron is measured within $\pm\Delta x$. Simultaneously, the x component of the momentum of that same electron is measured within $\pm\Delta p_x$. What can be concluded about the relationship between Δx and Δp_x?

(A) $\Delta x \geq h/\pi\Delta p_x$ (B) $\Delta x \geq h/2\pi\Delta p_x$
(C) $\Delta x \geq h/8\pi\Delta p_x$
(D) Nothing; the Heisenberg uncertainty principle does not apply here.

8. The position of an electron is measured within $\pm\Delta x$. Then the x component of the momentum of that same electron is measured within $\pm\Delta p_x$. What can be concluded about the relationship between Δx and Δp_x?

(A) $\Delta x \geq h/\pi\Delta p_x$ (B) $\Delta x \geq h/2\pi\Delta p_x$
(C) $\Delta x \geq h/8\pi\Delta p_x$
(D) Nothing; the Heisenberg uncertainty principle does not apply here.

9. The position of an electron is measured within $\pm \Delta x$. Simultaneously, the y component of the momentum of that same electron is measured within $\pm \Delta p_y$. What can be concluded about the relationship between Δx and Δp_y?

(A) $\Delta x \geq h/\pi\Delta p_y$ (B) $\Delta x \geq h/2\pi\Delta p_y$
(C) $\Delta x \geq h/8\pi\Delta p_y$
(D) Nothing; the Heisenberg uncertainty principle does not apply here.

46-5 The Wave Function

10. For the following questions use Eq. 46-11, $e^{i\theta} = \cos\theta + i\sin\theta$.

(*a*) What is θ if $e^{i\theta} = -1$?
(A) $\theta = 0$ (B) $\theta = -1$
(C) $\theta = \pi/2$ (D) $\theta = \pi$

(*b*) Which of the following is a solution to $\sqrt[3]{-1}$?
(A) $\cos\pi/2 + i\sin\pi/2$ (B) $\cos\pi/3 + i\sin\pi/3$
(C) $\cos\pi/6 + i\sin\pi/6$ (D) $\frac{1}{3}(\cos\pi/2 + i\sin\pi/2)$

46-6 Schrödinger's Equation

46-7 Barrier Tunneling

11. A particle with energy $E = U_0/2$ is incident on a barrier of height U_0 and thickness L. The transmission coefficient, under conditions for which Eq. 46-25 applies, is T_0. If the barrier thickness is doubled, with its height and the particle energy remaining unchanged, the new transmission coefficient will be

(A) $T_0^2/4$ (B) T_0^2 (C) $T_0/2$ (D) $T_0/4$

12. Protons and alpha particles, each with the same total energy, are incident on the same potential barrier. The transmission coefficient for the protons, under conditions for which Eq. 46-25 applies, is T_p. Compared with T_p, the transmission coefficient for alpha particles is

(A) much larger (B) slightly larger
(C) about equal (D) slightly smaller
(E) much smaller

QUESTIONS

1. How can the wavelength of an electron be given by $\lambda = h/p$? Does the very presence of the momentum p in this formula not imply that the electron is a particle?

2. In a repetition of Thomson's experiment for measuring e/m for the electron (see Section 32-2), a beam of electrons is collimated by passage through a slit. Why is the beam-like character of the emerging electrons not destroyed by diffraction of the electron wave at this slit?

3. Why is the wave nature of matter not more apparent in our daily observations?

4. Considering the wave behavior of electrons, we should expect to be able to construct an "electron microscope" using short-wavelength electrons to provide high resolution. This, indeed, has been done. (*a*) How might an electron beam be focused? (*b*) What advantages might an electron microscope have over a light microscope? (*c*) Why not make a proton microscope? A neutron microscope?

5. How many experiments can you recall that support the wave theory of light? The particle theory of light? The wave theory of matter? The particle theory of matter?

6. Is an electron a particle? Is it a wave? Explain your answer, citing relevant experimental evidence.

7. What common expression can be used for the momentum of either a photon or a particle?

8. Discuss the analogy between (*a*) wave optics and geometrical optics and (*b*) wave mechanics and classical mechanics.

9. Does a photon have a de Broglie wavelength? Explain.

10. Discuss similarities and differences between a matter wave and an electromagnetic wave.

11. If, in the de Broglie formula $\lambda = h/mv$, we let $m \rightarrow \infty$, do we get the classical result for particles of matter?

12. Considering electrons and photons as particles, how are they different from each other?

13. Is Eq. 46-1 for the de Broglie wavelength, $\lambda = h/p$, valid for a relativistic particle? Justify your answer.

14. How could Davisson and Germer be sure that the "54-V" peak of Fig. 46-2 was a first-order diffraction peak?

15. Do electron diffraction experiments give different information about crystals than can be obtained from x-ray diffraction experiments? From neutron diffraction experiments? Give examples.

16. Why are the hydrogen atoms clearly visible in Fig. 46-8 but not in Fig. 43-17?

17. In Fig. 46-3*b* (made with x rays) the diffraction circles are speckled, but in Fig. 46-3*c* (made with electrons) they are smooth. Can you explain why?

18. Electromagnetic waves will penetrate seawater to a certain extent if their frequency is low enough. This is the basis of one plan to communicate with submerged submarines. A difficulty with this plan is that the lower the frequency, the longer the time it takes to transmit a message (in Morse code pulses, say). Can you explain why this should be?

19. Why is the Heisenberg uncertainty principle not more readily apparent in our daily observations?

20. (*a*) Give examples of how the process of measurement disturbs the system being measured. (*b*) Can the disturbances be taken into account ahead of time by suitable calculations?

21. You measure the pressure in a tire, using a pressure gauge. The gauge, however, bleeds a little air from the tire in the process, so that the act of measuring changes the property that you are trying to measure. Is this an example of the Heisenberg uncertainty principle? Explain.

22. "The energy of the ground state of an atomic system can be precisely known, but the energies of its excited states are always subject to some uncertainty." Can you explain this statement on the basis of the uncertainty principle?

23. "If an electron is localized in space, its momentum becomes uncertain. If it is localized in time, its energy becomes uncertain." Explain this statement.

24. We have seen that barrier tunneling works for matter waves and for electromagnetic waves. Do you think that it also works for water waves? For sound waves?

25. Comment on the statement: "A particle cannot be detected while tunneling through a barrier, so that it does not make sense to say that such a thing actually happens."

26. List examples of barrier tunneling occurring in nature and in manufactured devices.

27. Several groups of experimenters are trying to detect gravity waves, perhaps coming from our galactic center, by measuring small distortions in a massive object through which the hypothesized waves pass. They seek to measure displacements as small as 10^{-21} m. (The radius of a proton is $\approx 10^{-15}$ m, a million times larger!) Does the uncertainty principle put any restriction on the precision with which this measurement can be carried out?

28. A proton and a deuteron, each having 3 MeV of energy, attempt to penetrate a rectangular potential barrier of height 10 MeV. Which particle has the higher probability of succeeding? Explain in qualitative terms.

EXERCISES

46-1 Matter Waves

1. A bullet of mass 41 g travels at 960 m/s. (*a*) What wavelength can we associate with it? (*b*) Why does the wave nature of the bullet not reveal itself through diffraction effects?

2. Using the classical relation between momentum and kinetic energy, show that the de Broglie wavelength of an electron can be written (*a*) as

$$\lambda = \frac{1.226\ \text{nm}}{\sqrt{K}}.$$

in which K is the kinetic energy in electron volts, or (*b*) as

$$\lambda = \sqrt{\frac{1.50\ \text{V}}{V}}\ \text{nm},$$

where V is the accelerating potential in volts. (Use the best values of the needed constants as found in Appendix B.)

3. Calculate the wavelength of a 1.00-keV (*a*) electron, (*b*) photon, and (*c*) neutron.

4. The wavelength of the yellow spectral emission line of sodium is 589 nm. At what kinetic energy would an electron have the same de Broglie wavelength?

5. If the de Broglie wavelength of a proton is 0.113 pm, (*a*) what is the speed of the proton and (*b*) through what electric potential difference would the proton have to be accelerated from rest to acquire this speed?

6. Singly charged sodium ions are accelerated through a potential difference of 325 V. (*a*) What is the momentum acquired by the ions? (*b*) Calculate their de Broglie wavelength.

7. The existence of the atomic nucleus was discovered in 1911 by Ernest Rutherford, who properly interpreted some experiments in which a beam of alpha particles was scattered from a foil of atoms such as gold. (*a*) If the alpha particles had a kinetic energy of 7.5 MeV, what was their de Broglie wavelength? (*b*) Should the wave nature of the incident alpha particles have been taken into account in interpreting these experiments? The distance of closest approach of the alpha particle to the nucleus in these experiments was about 30 fm. (The wave nature of matter was not postulated until more than a decade after these crucial experiments were first performed.)

8. The highest achievable resolving power of a microscope is limited only by the wavelength used; that is, the smallest detail that can be separated is about equal to the wavelength. Suppose one wishes to "see" inside an atom. Assuming the atom to have a diameter of 100 pm, this means that we wish to resolve detail of separation about 10 pm. (*a*) If an electron microscope is used, what minimum energy of electrons is needed? (*b*) If a light microscope is used, what minimum energy of photons is needed? (*c*) Which microscope seems more practical for this purpose? Why?

9. The 50-GeV electron accelerator at Stanford provides an electron beam of small wavelength that is suitable for probing the fine details of nuclear structure by scattering experiments. What is this wavelength and how does it compare with the size of an average nucleus? (At these energies it is sufficient to use the extreme relativistic relationship between momentum and energy; namely, $p = E/c$. This is the same relationship used for light and is justified when the kinetic energy of a particle is much greater than its rest energy, as in this case. The radius of a middle-mass nucleus is about 5.0 fm.)

10. Consider a balloon filled with (monatomic) helium gas at 18°C and 1.0 atm pressure. Calculate (*a*) the average de Broglie wavelength of the helium atoms and (*b*) the average distance between the atoms. Can the atoms be treated as particles under these conditions?

11. A nonrelativistic particle is moving three times as fast as an electron. The ratio of their de Broglie wavelengths, particle to electron, is 1.813×10^{-4}. By calculating its mass, identify the particle. See Appendix B.

12. (*a*) A photon in free space has an energy of 1.5 eV and an electron, also in free space, has a kinetic energy of that same amount. What are their wavelengths? (*b*) Repeat for an energy of 1.5 GeV.

13. In an ordinary color television set, electrons are accelerated through a potential difference of 25.0 kV. Find the de Broglie wavelength of such electrons (*a*) using the classical expression for momentum and (*b*) taking relativity into account.

14. What accelerating voltage would be required for electrons in an electron microscope to obtain the same ultimate resolving power as that which could be obtained from a gamma-ray microscope using 136-keV gamma rays? (Hint: See Exercise 8.)

46-2 Testing de Broglie's Hypothesis

15. A neutron crystal spectrometer utilizes crystal planes of spacing $d = 73.2$ pm in a beryllium crystal. What must be the Bragg angle θ so that only neutrons of energy $K = 4.2$ eV are reflected? Consider only first-order reflections.

16. A beam of thermal neutrons from a nuclear reactor falls on a crystal of calcium fluoride, the beam direction making an angle θ with the surface of the crystal. The atomic planes parallel to the crystal surface have an interplanar spacing of 54.64 pm. The de Broglie wavelength of neutrons in the incident beam is 11.00 pm. For what values of θ will the first three orders of Bragg-reflected neutron beams occur? (Hint: Neutrons, which carry no charge and are thus not subject to electrical forces, are not refracted as they pass through a crystal surface. Thus neutron diffraction can be treated in strict analogy with x-ray diffraction.)

17. In the experiment of Davisson and Germer (*a*) at what angles would the second- and third-order diffracted beams corresponding to a strong maximum in Fig. 46-2 occur, provided they are present? (*b*) At what angle would the first-order diffracted beam occur if the accelerating potential were changed from 54 to 60 V?

18. A potassium chloride (KCl) crystal is cut so that the layers of atomic planes parallel to its surface have a spacing of 314 pm between adjacent lines of atoms. A beam of 380-eV electrons is incident normally on the crystal surface. Calculate the angles ϕ at which the detector must be positioned to record strongly diffracted beams of all orders present.

46-3 Waves and Particles

19. Using a rotating shutter arrangement, you listen to a 540-Hz standard tuning fork for 0.23 s. What approximate spread of frequencies is contained in this acoustic pulse?

20. The signal from a television station contains pulses of full width $\Delta t \approx 10$ ns. Is it feasible to transmit television in the AM broadcasting band, which runs from about 500 to 1600 kHz?

46-4 Heisenberg's Uncertainty Principle

21. A nucleus in an excited state will return to its ground state, emitting a gamma ray in the process. If its mean lifetime is 8.7 ps in a particular excited state of energy 1.32 MeV, find the uncertainty in the energy of the corresponding emitted gamma-ray photon.

22. An atom in an excited state has a lifetime of 12 ns; in a second excited state the lifetime is 23 ns. What is the uncertainty in energy for a photon emitted when an electron makes a transition between these two states?

23. A microscope using photons is employed to locate an electron in an atom to within a distance of 12 pm. What is the minimum uncertainty in the momentum of the electron located in this way?

24. Imagine playing baseball in a universe where Planck's constant was 0.60 J · s. What would be the uncertainty in the position of a 0.50-kg baseball moving at 20 m/s with an uncertainty in velocity of 1.2 m/s? Why would it be hard to catch such a ball?

25. Find the uncertainty in the location of a particle, in terms of its de Broglie wavelength λ, so that the uncertainty in its velocity is equal to its velocity.

26. An electron is confined to a box the size of an atom so that one side of the box has length $d = 0.1$ nm. (*a*) Calculate the uncertainty in the momentum Δp of the electron. (*b*) Assuming that the electron is "bouncing" around inside the box with momentum $p \approx \Delta p$, calculate the kinetic energy of the electron. Express your answer in electron-volts.

27. Repeat Exercise 26 to find the kinetic energy of an electron confined to a nucleus, which has a size on the order of 10^{-14} m.

46-5 The Wave Function

46-6 Schrödinger's Equation

46-7 Barrier Tunneling

28. In Sample Problem 46-6, suppose that you can vary the thickness L of the barrier. To what value should the thickness be adjusted so that 1 electron out of 100 striking the barrier will tunnel through it?

29. (*a*) A proton and (*b*) a deuteron (which has the same charge as a proton but twice the mass) are incident on a barrier of thickness 10 fm and height 10 MeV. Each particle has a kinetic energy of 3.0 MeV. Find the transmission coefficients for them.

30. Suppose that an incident beam of 5.0-eV protons fell on a barrier of height 6.0 eV and thickness 0.70 nm, and at a rate equivalent to a current of 1.0 kA. How long would you have to wait—on the average—for one proton to be transmitted?

PROBLEMS

1. A beam of low-energy neutrons emerges from a reactor and is diffracted from a crystal. The kinetic energies of the neutrons are contained in a band of width ΔK centered on kinetic energy K. Show that the angles for a given order of diffraction are spread over a range $\Delta\theta$ given in degrees by

$$\Delta\theta = \left(\frac{90}{\pi}\right)(\tan\theta)\,\frac{\Delta K}{K},$$

where θ is the diffraction angle for a neutron with kinetic energy K.

2. A beam of atoms emerges from an oven that is at a temperature T. The distribution of the speeds of the atoms in the beam is proportional to $v^3 e^{-mv^2/2kT}$ (see Section 22-4). (*a*) Show that the distribution of de Broglie wavelengths of the atoms is proportional to $\lambda^{-5} e^{-h^2/2mkT\lambda^2}$, and (*b*) that the most probable de Broglie wavelength is

$$\lambda_p = \frac{h}{\sqrt{5mkT}}.$$

3. Consider a barrier such as that of Fig. 46-16, whose height U_0 is 6.00 eV and whose thickness L is 700 pm. Calculate the energy of an incident electron such that its transmission coefficient is 1 in 1000.

4. Consider the barrier-tunneling situation defined by Sample Problem 46-6. What fractional change in the transmission coefficient occurs for a 1% increase in (*a*) the barrier height, (*b*) the barrier thickness, and (*c*) the incident energy of the electron?

5. Use Eq. 46-12 to verify the trig identity $\cos(a + b) = \cos a \cos b - \sin a \sin b$.

6. Particles of mass m and energy E originally traveling to the right in a region of space where $U = 0$ strike a *step potential* at $x = 0$ of height $U_0 < E$ for all $x > 0$. Classically all particles will pass over the potential step. Solve the quantum mechanics problem by following a process similar to that described in Section 46-7. (*a*) Write down solutions valid in the region $x < 0$ consisting of an incident wave of amplitude ψ_1 and a reflected wave of amplitude ψ_2, and in the region $x > 0$ consisting of a transmitted wave of amplitude ψ_3. Express the solutions in terms of $k = (2\pi/h)\sqrt{2mE}$ and $k' = (2\pi/h)\sqrt{2m(E - U_0)}$. (*b*) For a mathematically well-behaved wave, the wave function and its first derivative must be continuous across the boundary; that is, at $x = 0$ we must have $\psi_{x<0} = \psi_{x>0}$ and $d\psi_{x<0}/dx = d\psi_{x>0}/dx$. Apply these *boundary conditions* to find ψ_2/ψ_1 and ψ_3/ψ_1. (*c*) Show that the transmission coefficient, in this case defined as $|\psi_3|^2/|\psi_1|^2$, is

$$T = 4(1 + k'/k)^{-2}.$$

COMPUTER PROBLEM

1. A ping-pong ball can be dropped from a height of 1 meter directly above a bowling ball. The ping-pong ball will fall, strike the bowling ball, bounce back up, fall back down, and so on, always perfectly elastically. If the ping-pong ball is dropped exactly above the center of the bowling ball, it will continue to bounce forever. Heisenberg's uncertainty principle limits the precision with which the ball can be dropped, so that it will eventually bounce off the bowling ball. (a) Write a computer program to model the bouncing of the ping-pong ball assuming that it is dropped a horizontal distance Δx from directly above the center of the bowling ball with an initial horizontal momentum Δp_x. (b) Assuming that Δx and Δp_x are related by Heisenberg's uncertainty principle, find the maximum number of times the ping-pong ball can bounce from the bowling ball.

CHAPTER 47

ELECTRONS IN POTENTIAL WELLS

In the previous chapter we studied the motions of free electrons. Here we study the motions of bound electrons*—that is, electrons that are not free to roam beyond certain limits.*

Although the energy of a completely free electron is not *quantized, we will learn that the energy of a bound electron* is *quantized. Our principal interest is the motions of electrons bound into atoms. As a preliminary, we study electrons bound into artificial structures called* potential wells.

We then focus our attention on the single electron of the hydrogen atom, dealing first with the model of the hydrogen atom proposed by Bohr at a time when quantum mechanics was not yet fully developed. We then turn to the modern quantum treatment of the hydrogen atom. We identify the quantum numbers that specify its quantum states and we explore the behavior of the electron in a few of these states.

47-1 ELECTRONS, FREE AND BOUND

In Chapter 46 we studied the motion of a free electron—that is, an electron on which no net force acts. In this chapter we study the motion of a *bound electron*—that is, an electron on which forces act to confine it to a restricted region of space, perhaps, within the volume of an atom. One difference between a free electron and an electron bound in an atom immediately springs to mind. A free electron moves in only one dimension, whereas an electron bound into an atom can move in three dimensions.

However, the principal difference between a free and a bound electron is that, as we shall see in the next section, the very fact that a bound electron is *localized* means that its energy can have only values belonging to a certain discrete set. That is, its energy is *quantized.* The energy of a free electron is not quantized.

To separate the effects of localization and dimensionality, we study first an electron moving in a somewhat artificial "atom" in which the electron is localized but the motion continues to be one-dimensional.

47-2 AN ELECTRON TRAPPED IN A POTENTIAL WELL

Figure 47-1*a* shows an arrangement that might be adapted to localize an electron—in this case, to confine it to one-dimensional motion in the central region of length *L*. Figure 47-1*b* shows $U(x)$, the variation with x of the electron's potential energy in such an arrangement. We call such a potential energy variation a potential energy well or, for short, a *potential well.* In practice, electromechanical devices to trap charged particles must contain features not shown in Fig. 47-1*a*. Also, it is possible to construct devices on the scale of atomic dimensions that achieve the same result. Such *quantum well* structures are routinely fabricated in the laboratory.

Let $E\,(=K + U)$ in Fig. 47-1*b* represent the total energy of the electron. We take $U(x) = 0$ within the well, so that the total energy E of the electron is entirely kinetic in this region. When the electron reaches either end of the central region, it encounters a narrow zone in which its potential energy changes rapidly from zero to U_0. At either end of the central region, then, a force acts on the electron, tending to decrease its speed and turn it around. According

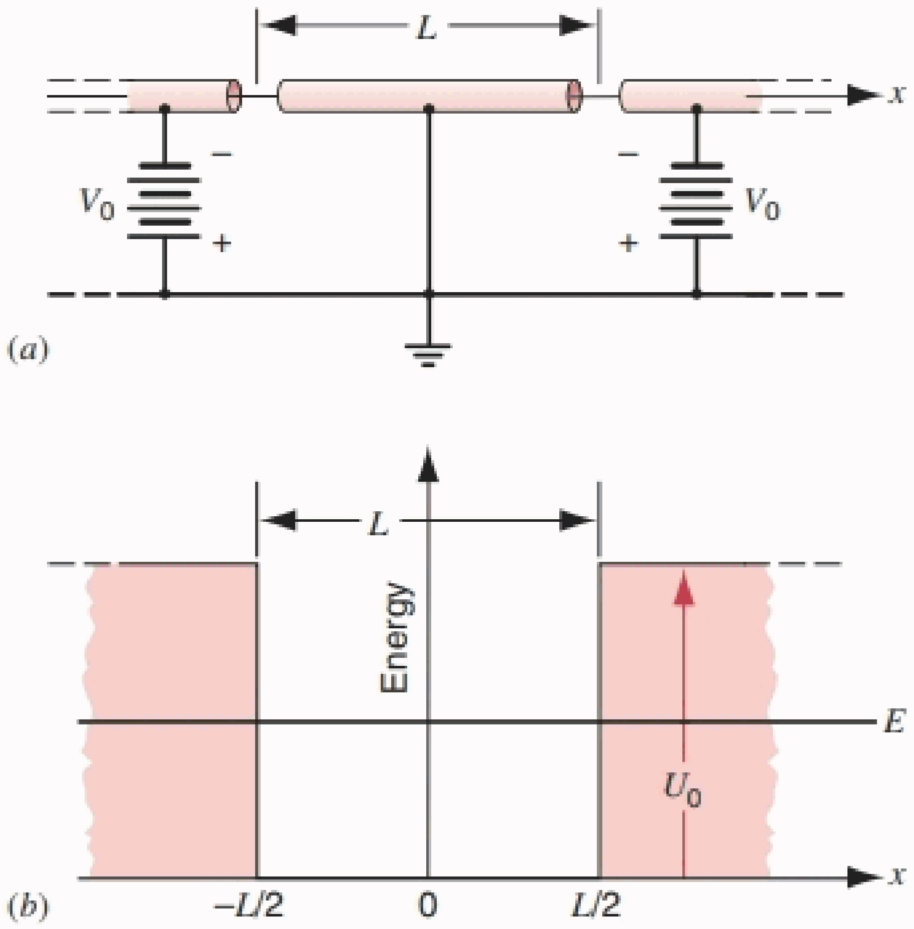


Figure 47-1. (*a*) An arrangement of hollow metal tubes that might be used to confine an electron to a central region of length *L*. (*b*) The potential energy of an electron of total energy *E* in the above arrangement. The classically forbidden regions of the potential well, whose depth is U_0, are shaded. The origin of the *x* axis is taken to be the center of the well.

to Newtonian mechanics, the electron can only bounce back and forth in the central region between turning points at $x = -L/2$ and $x = +L/2$. The electron's actual behavior, however, is not described by Newtonian mechanics but by quantum mechanics.

An Infinitely Deep Potential Well

We now consider the special case in which the heights U_0 of the barriers between which the electron of Fig. 47-1*b* is bound are infinite, as Fig. 47-2 suggests. We call such a structure an infinitely deep potential energy well or, for short, an *infinite well.*

In Section 46-7 we saw that, although an electron can penetrate a barrier placed in its path, it cannot do so if the barrier is infinitely tall. In that case the electron, rebounding from the barrier like a tennis ball from a brick wall, behaves as if Newtonian physics prevailed. A similar situation holds for the infinite well of Fig. 47-2. The electron trapped in the well cannot penetrate the walls of the well but can only rattle back and forth between those walls. In quantum terms, $\psi(x)$, the space-varying part of the wave function that describes the electron bound in an infinite well, must have the value zero at the walls and at all points beyond the walls, signifying that the probability of finding the electron in those locations is zero.

The fact that $\psi(x)$ must vanish at the walls reminds us of the behavior of a string stretched between two rigid supports. In this case, standing waves can be set up in the string, subject to the *boundary condition* that the displacement of the string is zero at the two rigid supports (see Fig. 18-20). We can ease our introduction to quantum mechanics by exploring the analogy between mechanical waves propagating along a stretched string and matter waves associated with an electron trapped in an infinite well.

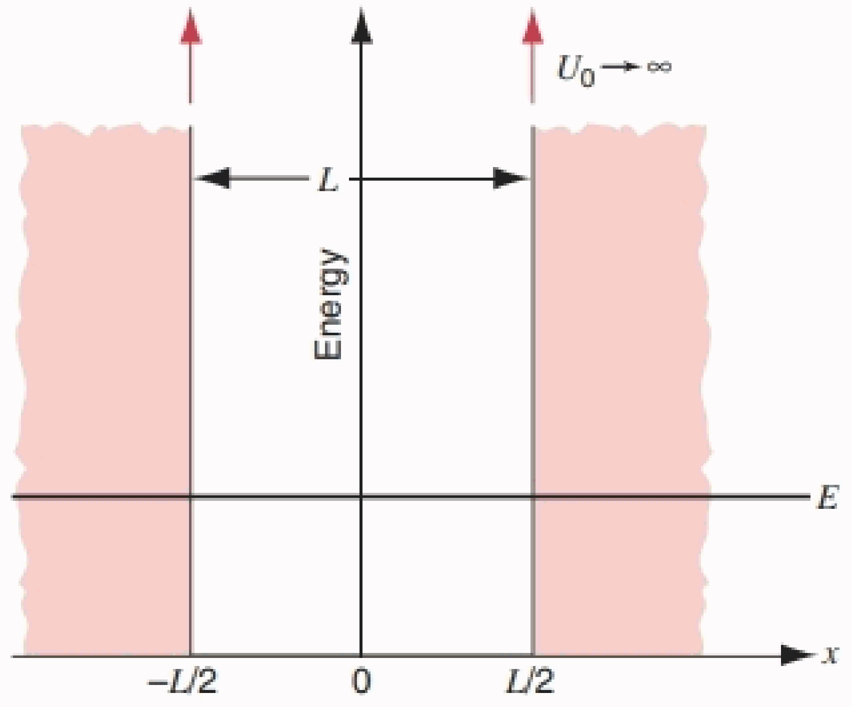


Figure 47-2. An infinite well. Compare with the well of Fig. 47-1*b*, whose depth is finite. The classically forbidden regions are shaded.

Mechanical waves of *any* wavelength can be propagated along a stretched string if the string is long enough so that its length may be considered infinite. However, only waves having a certain discrete set of wavelengths can be propagated along a stretched string of finite length, as in Fig. 18-20. We may say that, once the length of the string is taken to be finite—that is, once the waves on the string are *localized*—the wavelengths of these permitted waves become *quantized.*

When considering matter waves, it is customary to think in terms of the energy of the associated particle rather than the wavelength of the matter wave itself. We have seen in Chapter 46 that a free particle—that is, a particle not acted on by any force—can move with *any* reasonable value of the energy *E*. We will see below that, if the particle is not free but is localized to a finite region of space, its energy can take on only values belonging to a discrete set. That is, once the electron is *localized* to a finite region of space, its energy becomes *quantized.* Thus, for both mechanical waves on a stretched string and for matter waves,

> *Localization leads to quantization.*

The analogy between matter waves and mechanical waves has led us to conclude that electrons confined to an infinite well can possess only energies that belong to a discrete set. Let us now find these energies.

Energy Levels for an Infinite Well

To describe the matter wave associated with an electron, we must find its (space-varying) wave function $\psi(x)$. The straightforward way to do so is to solve Schrödinger's equation for the conditions of the problem at hand. However, for the potential variation associated with an infinite well, it turns out that Schrödinger's equation is identical in mathematical form to the equation that describes the space variation of mechanical waves in a stretched string. Indeed, that is why—so far—we have found the analogy between matter waves and mechanical waves so useful.

Thus we can assume that $\psi(x)$, which describes the matter wave associated with an electron in an infinite well, has the same form as $y(x)$, which describes the displacement associated with a mechanical wave on a string of length L stretched between rigid supports. That is, we assume that only those matter waves can be present whose de Broglie wavelengths are such that an integral number of half-waves fit into the space L, or,

$$n\left(\frac{\lambda_n}{2}\right) = L \qquad (n = 1, 2, 3, \ldots),$$

in which we call the integer n a *quantum number.* Solving for λ_n yields

$$\lambda_n = \frac{2L}{n} \qquad (n = 1, 2, 3, \ldots), \qquad (47\text{-}1)$$

which identifies the discrete set of de Broglie wavelengths that can exist in the well.

The energy of the trapped electron is entirely kinetic because (see Fig. 47-2) we have taken $U(x) = 0$ inside the well. The magnitude of the allowed values of the electron's momentum is given, using Eq. 46-1 ($\lambda = h/p$) and Eq. 47-1, by

$$p_n = \frac{h}{\lambda_n} = \frac{nh}{2L} \qquad (n = 1, 2, 3, \ldots). \qquad (47\text{-}2)$$

Knowing the momenta of the allowed quantum states from Eq. 47-2, we can calculate the corresponding energies from

$$E_n = \frac{p_n^2}{2m} = n^2\frac{h^2}{8mL^2} \qquad (n = 1, 2, 3, \ldots). \qquad (47\text{-}3)$$

Equation 47-3 defines the energies of the allowed *quantum states* for an electron trapped in an infinite well of width L. Figure 47-3 shows several *energy levels,* as we call them, for an infinite well with $L = 100$ pm, a number chosen because it is approximately the size of a typical atom. The lowest energy level is called the *ground state,* and the states above the ground state are called *excited states.*

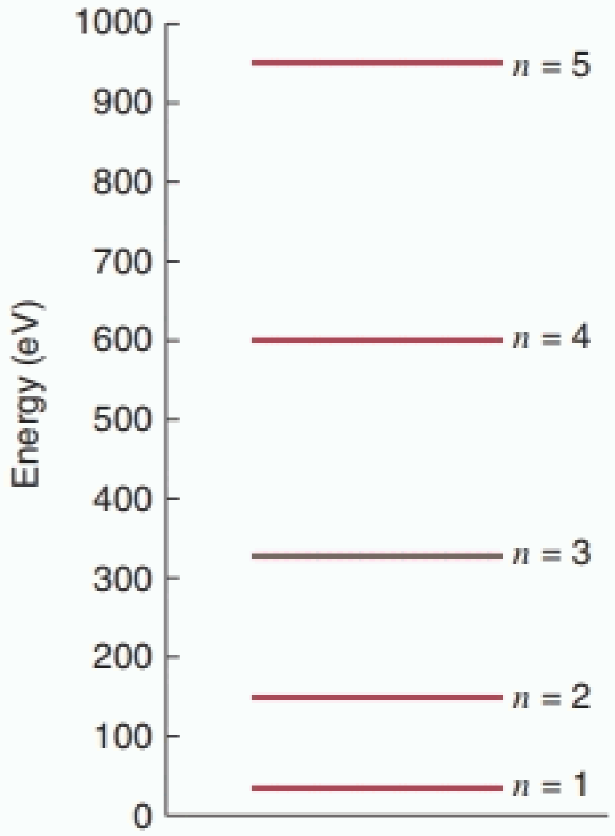


FIGURE 47-3. The energy levels of the five quantum states whose quantum numbers are shown. The energies were calculated for an infinite well of width $L = 100$ pm, using Eq. 47-3.

Probability Densities for an Infinite Well

It is usually more convenient to deal with *wave numbers* than with wavelengths. By the definition of wave number, Eq. 47-1 then becomes

$$k_n = \frac{2\pi}{\lambda_n} = \frac{n\pi}{L} \qquad (n = 1, 2, 3, \ldots). \qquad (47\text{-}4)$$

By analogy to the allowed mechanical waves in a stretched string of length L, the wave function can be expressed in terms of sines and cosines. If an *odd* number of half-waves are to fit into the well, the wave function can be written as

$$\psi_n(x) = \psi_{\text{odd}} \cos k_n x$$
$$= \psi_{\text{odd}} \cos\frac{n\pi x}{L} \qquad (n = 1, 3, 5, \ldots), \qquad (47\text{-}5)$$

in which ψ_{odd}, the amplitude of the matter wave, is an adjustable constant. Convince yourself that this wave function satisfies the condition that $\psi(x) = 0$ at the walls ($x = \pm L/2$) for odd values of n.

Similarly, if an *even* number of half-waves are to fit into the well, the wave function is given by

$$\psi_n(x) = \psi_{\text{even}} \sin k_n x$$
$$= \psi_{\text{even}} \sin\frac{n\pi x}{L} \qquad (n = 2, 4, 6, \ldots), \qquad (47\text{-}6)$$

in which ψ_{even} is also an amplitude constant. For this wave function, $\psi(x) = 0$ at the walls for even values of n. Figure 47-4*a* shows $\psi(x)$ for the quantum numbers $n = 1$, $n = 2$, and $n = 3$.

Just as for a free particle, the quantity $\psi\psi^*$ gives the *probability density* $P(x)$ of the matter wave. From Eqs. 47-5 and 47-6 we then have

$$P_n(x) = \psi\psi^* = \psi_{\text{odd}}^2 \cos^2\frac{n\pi x}{L} \qquad (n = 1, 3, 5, \ldots) \qquad (47\text{-}7)$$

and

$$P_n(x) = \psi\psi^* = \psi_{\text{even}}^2 \sin^2\frac{n\pi x}{L} \qquad (n = 2, 4, 6, \ldots). \qquad (47\text{-}8)$$

Note that $n = 0$ in Eq. 47-8 is not an option, because that would lead to $P_0(x) = 0$, which tells us that the electron is not present anywhere within the well. The well would be empty!

The three probability distributions shown in Fig. 47-4*b* show very different states of motion for the electron. For example, $P_1(x)$ shows that an electron in a state with energy E_1 is most likely to be found near $x = 0$ (in the center of the well). An electron in the $n = 2$ state, on the other hand, is never found at the center of the well; you should look for it near $x = \pm L/4$, where it is most likely to be found.

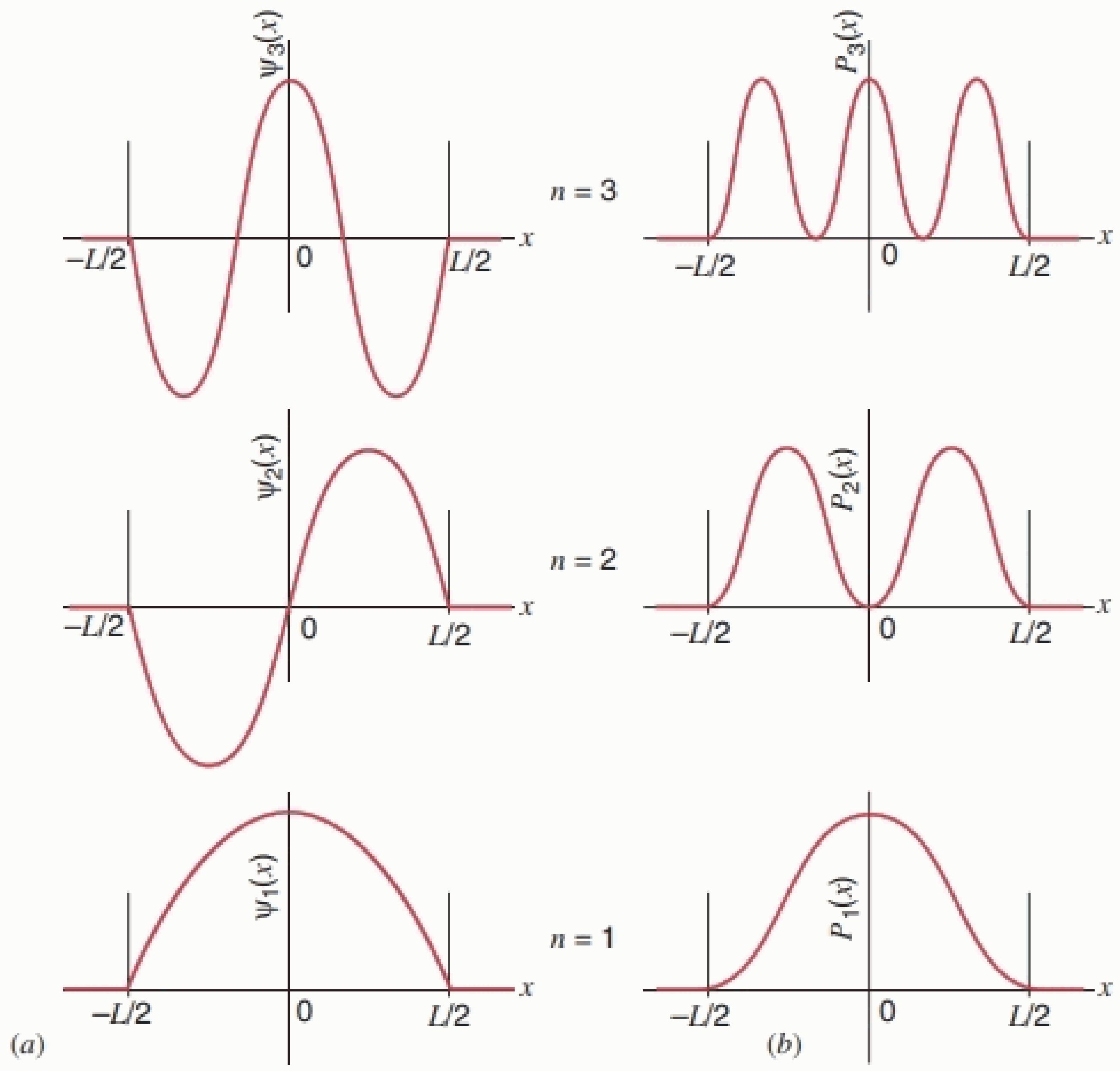


FIGURE 47-4. (*a*) The space-dependent part of the wave function $\psi_n(x)$ for three quantum states, with their quantum numbers indicated. (*b*) The probability density $P_n(x)$, computed from Eqs. 47-7 and 47-8, for the same three quantum states.

Figure 47-4*b* suggests that, for quantum states with increasingly higher quantum numbers, the trapped electron is more and more likely to be found with equal probability in all parts of the well. Figure 47-5 shows the probability density for $n = 15$. Adjacent peaks are separated by $\Delta x = L/15$. Note that the probability of locating the electron within a few multiples of this basic interval is the same for all parts of the well, much as it would be if Newtonian physics prevailed. (In Newtonian physics, the probability of locating the particle in any interval in the well does not depend on the location of the interval.) Consider what a plot similar to Fig. 47-5 would look like for a really high quantum number—say, $n = 5000$! This is an example of the *correspondence principle,* which we can state as follows:

> *At increasingly high quantum numbers, the predictions of quantum mechanics merge smoothly with those of Newtonian mechanics.*

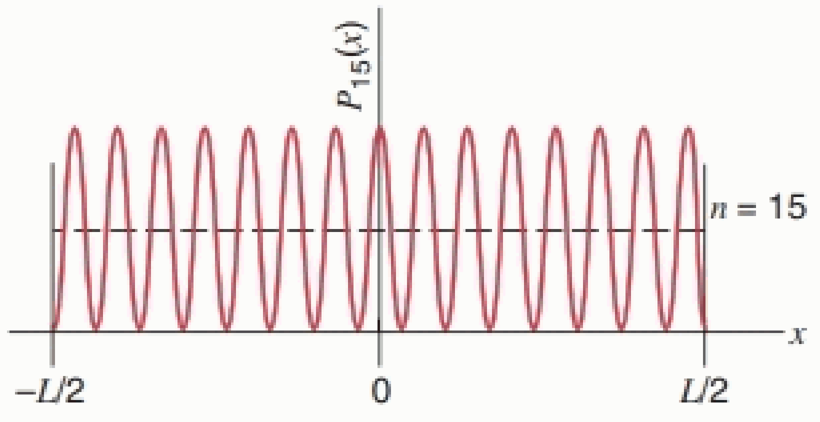


FIGURE 47-5. The probability density for an electron trapped in an infinite well of width L for the quantum state with $n = 15$. The dashed line represents the classical uniform probability distribution.

Normalization

We cannot give numerical values to the probability density $P_n(x)$ until we have evaluated the amplitude constants ψ^2_{odd} in Eq. 47-7 and ψ^2_{even} in Eq. 47-8. We do so by realizing that, whatever its quantum state, the electron must be *somewhere* within the infinite well of Fig. 47-2. That is, recalling from Section 46-5 that $P(x)\,dx$ $[= \psi^2(x)\,dx]$ gives the probability of finding the electron between x and $x + dx$, the sum of the probabilities in all such intervals in the well must give 100% or unity. In equation form,

$$\int_{-L/2}^{L/2} P_n(x)\,dx = 1. \qquad (47\text{-}9)$$

In Sample Problem 47-3 we will use this *normalization equation,* as it is called, to show that $\psi^2_{\text{odd}} = \psi^2_{\text{even}} = 2/L$, independent of the value of the quantum number n.

Zero-Point Energy

The energy of the lowest-energy quantum state (ground state) for an electron bound in an infinite well corresponds to $n = 1$ in Eq. 47-3. Equation 47-3 shows that the energy of this state is not zero but $h^2/8mL^2$. This minimum energy is called the *zero-point energy* for a particle of mass m bound in an infinite well of width L. This phenomenon exists in all quantum systems. Even at the absolute zero of temperature, where a particle is in its lowest possible energy state, the particle still has energy and momentum.

Heisenberg's uncertainty principle helps us to understand the existence of a zero-point energy. If a particle truly had zero energy ($E = 0$), it would also have zero momentum ($p_x = 0$). If p_x is known to be zero, then Δp_x, the uncertainty in p_x, must also be zero. However, if $\Delta p_x = 0$, the uncertainty principle $\Delta p_x \cdot \Delta x \geq h/2\pi$ (Eq. 46-8) requires that $\Delta x \to \infty$. Because Δx cannot be greater than L, the width of the well, we conclude that our initial assumption is wrong; p_x can not be zero, and a zero-point energy must exist.

SAMPLE PROBLEM 47-1. An electron is confined to an infinite well whose width L (= 100 pm) is roughly the size of an atom. (*a*) What are the energies of the four least-energetic quantum states? (*b*) What energy must be imparted to the electron to raise it from a state with $n = 12$ to a higher-energy state, with $n = 25$?

Solution (*a*) From Eq. 47-3, with $n = 1$ and, expressing the constants involved to four significant figures (see Appendix B), we find

$$\begin{aligned} E_1 &= n^2 \frac{h^2}{8mL^2} \\ &= (1)^2 \frac{(6.626 \times 10^{-34}\ \mathrm{J \cdot s})^2}{(8)(9.109 \times 10^{-31}\ \mathrm{kg})(100 \times 10^{-12}\ \mathrm{m})^2} \\ &= 6.025 \times 10^{-18}\ \mathrm{J} = 37.6\ \mathrm{eV}. \end{aligned}$$

To find the energies of the higher states, we write Eq. 47-3 as $E_n = n^2 E_1$, and so the energies of the quantum states with $n = 2$, $n = 3$, and $n = 4$ are

$$\begin{aligned} E_2 &= (2)^2(37.6\ \mathrm{eV}) = 150\ \mathrm{eV}, \\ E_3 &= (3)^2(37.6\ \mathrm{eV}) = 338\ \mathrm{eV}, \\ E_4 &= (4)^2(37.6\ \mathrm{eV}) = 602\ \mathrm{eV}. \end{aligned}$$

(*b*) The energy difference, from Eq. 47-3, is

$$\begin{aligned} \Delta E = E_{25} - E_{12} &= [(25)^2 - (12)^2]\,(37.6\ \mathrm{eV}) \\ &= 18{,}100\ \mathrm{eV} = 18.1\ \mathrm{keV}. \end{aligned}$$

SAMPLE PROBLEM 47-2. A 1-ng ($= 10^{-9}$ g $= 10^{-12}$ kg) speck of dust moves back and forth between two rigid walls separated by $L = 0.1$ mm. The particle moves so slowly that it takes 100 s for it to cross this small gap. What quantum number describes this motion?

Solution The energy of the particle is entirely kinetic and is given by

$$\begin{aligned} E &= \frac{1}{2} mv^2 = \frac{1}{2}(1 \times 10^{-12}\ \mathrm{kg})(1 \times 10^{-6}\ \mathrm{m/s})^2 \\ &= 5 \times 10^{-25}\ \mathrm{J}. \end{aligned}$$

Solving Eq. 47-3 for the quantum number n yields

$$\begin{aligned} n &= \frac{L}{h}\sqrt{8mE} \\ &= \frac{1 \times 10^{-4}\ \mathrm{m}}{6.63 \times 10^{-34}\ \mathrm{J \cdot s}} \sqrt{(8)(10^{-12}\ \mathrm{kg})(5 \times 10^{-25}\ \mathrm{J})} \\ &= 3 \times 10^{11}. \end{aligned}$$

This is a very large number. It is experimentally impossible to measure the difference between the energy of a quantum state with $n = 300{,}000{,}000{,}000$ and that of its two adjoining states, which have $n = 300{,}000{,}000{,}001$ and $n = 299{,}999{,}999{,}999$. The quantized nature of the motion in this example would never reveal itself.

If you compare this example with the previous one, you will see that, although its mass and its kinetic energy are both extremely small by "ordinary" standards, our speck of dust is *still* a gross macroscopic object when compared with an electron.

SAMPLE PROBLEM 47-3. Evaluate the normalization constant ψ^2_{odd} in Eq. 47-7, which gives (for odd quantum numbers) the probability density for a particle of mass m trapped in an infinite well of width L.

Solution Substituting Eq. 47-7 into the normalization equation, Eq. 47-9, yields

$$\psi^2_{\mathrm{odd}} \int_{-L/2}^{L/2} \cos^2 \frac{n\pi x}{L}\, dx = 1. \tag{47-10}$$

This integral can be carried out most easily by introducing a new variable θ, defined from

$$\theta = \frac{n\pi x}{L},$$

from which we see that

$$dx = \frac{L}{n\pi}\, d\theta.$$

The limits on the integral ($x = \pm L/2$) become the limits on the new variable θ, or $\theta = \pm n\pi/2$. With these changes, Eq. 47-10 becomes

$$\frac{\psi^2_{\mathrm{odd}} L}{n\pi} \int_{-n\pi/2}^{n\pi/2} \cos^2\theta\, d\theta = \frac{\psi^2_{\mathrm{odd}} L}{n\pi} \left(\frac{1}{2}\theta + \frac{1}{4}\sin 2\theta \right) \Bigg|_{-n\pi/2}^{n\pi/2} = 1$$

(The integral of $\cos^2\theta\, d\theta$ can be found in Appendix I.) Substituting the limits and solving for ψ^2_{odd} leads eventually to

$$\psi^2_{\mathrm{odd}} = \frac{2}{L}. \tag{47-11}$$

Note that the normalization constant (for odd n) does not depend on the quantum number n, the mass m, or the energy E of the trapped particles. If you repeat the above derivation for the case of even quantum numbers, you will find, using Eq. 47-8 in place of Eq. 47-7, that ψ^2_{even} is also equal to $2/L$. Thus $\psi^2_{\mathrm{odd}} = \psi^2_{\mathrm{even}} = 2/L$. Note that each of these quantities has the dimension of $(\mathrm{length})^{-1}$. It then follows from Eqs. 47-7 and 47-8 that, because dx is a length, the quantity $P_n(x)\, dx$ is dimensionless, which it must be if this product is to represent a probability.

SAMPLE PROBLEM 47-4. An electron is trapped in an infinitely deep well of width L. If the electron is in its ground state, what is the relative probability of finding the electron in the central third of the well?

Solution The probability density for the ground state, which corresponds to $n = 1$, is given from Eq. 47-7 as

$$P_1(x) = \frac{2}{L} \cos^2 \frac{\pi x}{L}.$$

The integral of $P_1(x)$ over the entire well is unity, signifying the certainty that the electron is *somewhere* within the well. The central third of the well extends from $x = -L/6$ to $x = +L/6$. Thus the fraction that we seek is

$$f = \int_{-L/6}^{L/6} P_1(x)\,dx = \frac{2}{L}\int_{-L/6}^{L/6} \cos^2\frac{\pi x}{L}\,dx.$$

Evaluating this integral as we did in Sample Problem 47-3 leads eventually to $f = 0.61$, regardless of the value of L. Thus the probability that the electron will be found in the central third of the infinite well is 61%. The probability that we will detect it in either of the outer thirds of the well is 19.5% ($0.195 + 0.61 + 0.195 = 1$). If the electron obeyed Newtonian physics (which it does not), we would be equally likely to find it in each of these three regions of the well.

47-3 AN ELECTRON TRAPPED IN A FINITE WELL

We now analyze the motion of an electron trapped in a well of finite, rather than infinite, depth. Figure 47-6*a* (see also Fig. 47-1*b*) shows such a well.

We can learn something by comparing a *finite well* and a *finite barrier.* In Section 46-7 we saw that an electron of total energy E approaching a potential barrier of finite width L and finite height U_0 (where $U_0 > E$) has a definite probability of appearing on the other side of the barrier even though, according to Newtonian mechanics, it cannot do so. This suggests that an electron trapped in the finite well of Fig. 47-6*a* can penetrate beyond the boundaries of the well even though, again according to Newtonian mechanics, it cannot be found there (because its kinetic energy $K = E - U$ would be negative).

The analogy to mechanical waves in a stretched string—so useful in analyzing an infinite well—fails us in the case of a finite well. We must find $\psi(x)$ in this case by solving Schrödinger's equation, which (see Eq. 46-17) is

$$-\frac{h^2}{8\pi^2 m}\frac{d^2\psi(x)}{dx^2} + U(x)\,\psi(x) = E\,\psi(x). \quad (47\text{-}12)$$

Here m is the mass of the electron, E is its total energy, and $U(x)$ is its potential energy. We find $\psi(x)$ for the finite well by solving Eq. 47-12 separately for the three regions of interest shown in Fig. 47-6*a*. Thus

Region 1: This region, for which $x \le -L/2$, is *outside* the well, on the left. Here $U(x) = U_0 =$ a positive constant, and $U(x) > E$.

Region 2: This region, for which $-L/2 \le x \le +L/2$, is *inside* the well, where we have taken $U(x) = 0$.

Region 3: This region, for which $x \ge +L/2$, is *outside* the well, on the right. Here $U(x) = U_0 =$ a positive constant and $U(x) > E$.

Solving Schrödinger's equation separately for each region yields three separate expressions for $\psi(x)$. We must then "stitch together" these three expressions subject to these three "commonsense" quantum requirements:

1. $\psi(x)$ must be finite everywhere. If it were not, the probability density $P(x)$ would be infinite at one or more points.

2. $\psi(x)$ must be continuous at $x = \pm L/2$. If it were not, the probability density would be undetermined at those points.

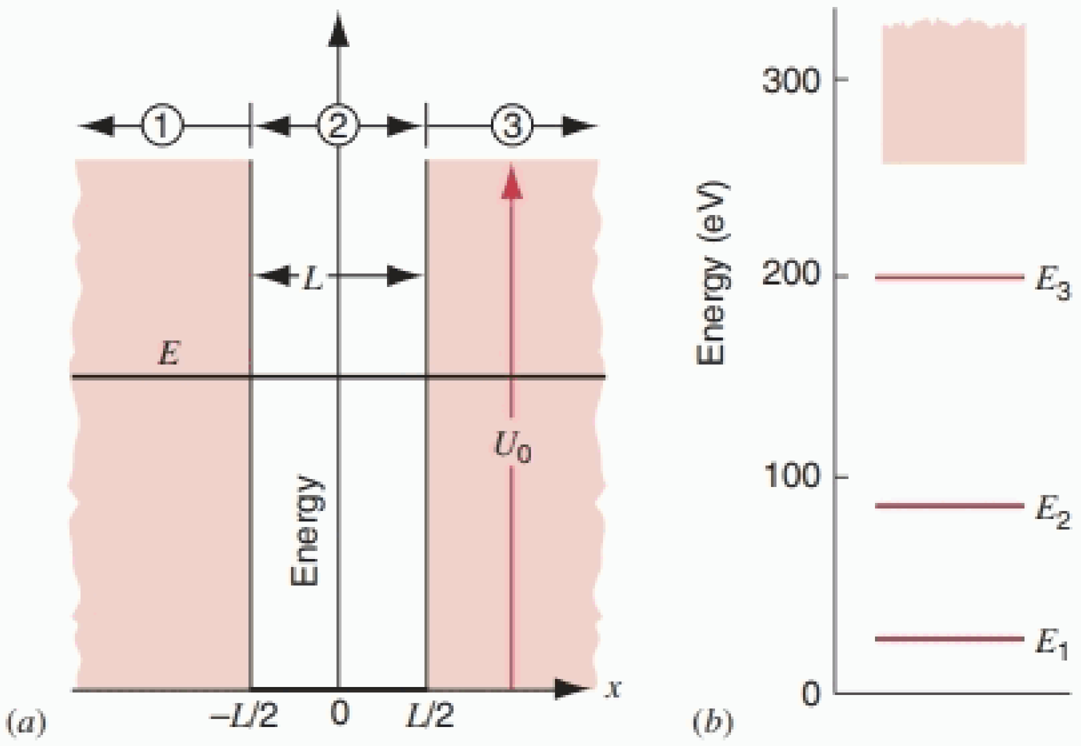


FIGURE 47-6. (*a*) A finite potential well of width L and depth U_0. (*b*) The three energy levels that can exist in this well for $L = 100$ pm and $U_0 = 250$ eV. The shaded area above $E = 250$ eV represents the smooth continuum of energies that can be possessed by electrons that are not trapped in the well.

3. The first derivative, $d\psi/dx$, must also be continuous at $x = \pm L/2$. If it were not, the quantity $d^2\psi/dx^2$ in Eq. 47-12 would be undetermined at those points.

To sum up, the resultant "stitched together" wave function for the finite well must not be infinite anywhere and must not have any "jumps" or "kinks."

Here are the solutions in the three regions. You should verify each of them by direct substitution into Eq. 47-12.

Region 1: Here we find

$$\psi_1(x) = Ae^{+k'x}, \quad (47\text{-}13)$$

where k', a positive quantity, is given by

$$k' = \frac{2\pi\sqrt{2m(U_0 - E)}}{h} \quad (47\text{-}14)$$

and A is an arbitrary constant. The function $\psi_1(x) = e^{-k'x}$ is also a formal solution in region 1, but we reject it because $\psi(x)$ becomes infinitely large as x approaches large negative values, which is a violation of requirement 1.

Region 2: Here we find

$$\psi_2(x) = B\sin kx + C\cos kx, \quad (47\text{-}15)$$

in which B and C are arbitrary constants and k, a positive quantity, is given by

$$k = \frac{2\pi\sqrt{2mE}}{h}. \quad (47\text{-}16)$$

Region 3: Here we find

$$\psi_3(x) = De^{-k'x}, \quad (47\text{-}17)$$

in which D is an arbitrary constant and k' is given by Eq. 47-14. The function $\psi_3(x) = e^{+k'x}$ is also a formal solution, but we reject it because it becomes infinitely large as x approaches large positive values.

It now remains to adjust the arbitrary constants A, B, C, and D that appear in Eqs. 47-13, 47-15, and 47-17, because we require that there are no "jumps" or "kinks" at the walls of the well. It is beyond our scope to show it in this text, but this program can be carried out *only for selected values of the total energy E.* It is precisely at this point that we learn that, because the electron is localized (it is confined to the well), its energy must be quantized.

Figure 47-6*b* shows the three energy levels that fit into the well of Fig. 47-6*a*, to which we have arbitrarily assigned the properties $L = 100$ pm and $U_0 = 250$ eV. Above the well—that is, for $E > U_0$—the shading suggests that the energy is no longer quantized. We refer to this region as *the continuum.* Here the energy of the electron can have any value.

Probability Densities for an Electron in a Finite Well

Figure 47-7 shows the probability densities for the three quantum states of an electron trapped in a finite well of width $L = 100$ ppm and depth $U_0 = 250$ eV. The wave functions have been *normalized* (see Eq. 47-9), which means that the total area under each curve, including all three regions of interest, is unity. The electron must be *somewhere*—if not inside the well, then outside.

We now call attention to three properties of the quantum states of an electron trapped in a finite well.

First property: As it happens, the three states shown in Fig. 47-6*b* are the only ones that can exist in this particular well. It can be shown that a fourth quantum state, corresponding to $n = 4$, can exist in our well if, keeping the width of the well at $L = 100$ ppm, we increase its depth from its present value of $U_0 = 250$ eV to a value somewhat above 340 eV. It is reasonable that a deeper well can accommodate more quantum states because the deeper the well the more it resembles an infinite well, which can accommodate *all* states, regardless of their quantum number.

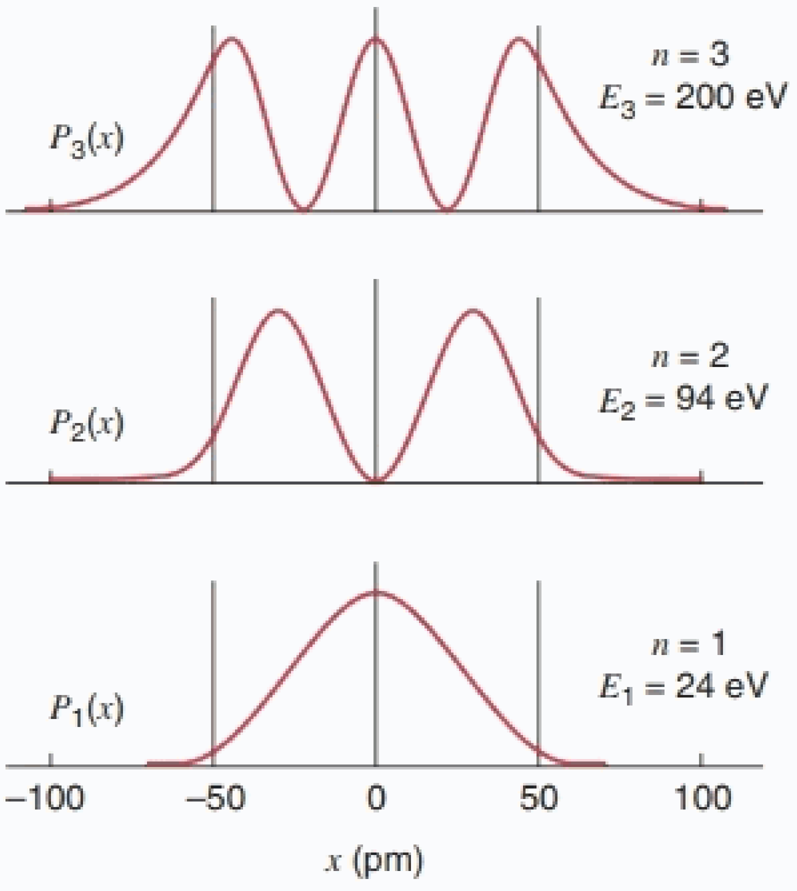


FIGURE 47-7. The probability densities $P_n(x)$, the energies, and the quantum numbers for an electron trapped in a finite well like that of Fig. 47-6*a*. The depth U_0 of the well is 250 eV and its width L is 100 pm. No other bound-electron quantum states can exist for this well.

Alternatively, we can accommodate the $n = 4$ quantum state in our well if, keeping the *depth* of the well at $U_0 = 250$ eV, we increase its width from its present value of $L = 100$ pm to a value somewhat above 115 pm. It is reasonable that making the well wider should permit it to accommodate additional quantum states because the wider the well, the more the particle trapped within it resembles a free particle, which can possess any energy.

Second property: As Fig. 47-7 shows, the probability of finding the electron outside the well increases as the quantum number increases. Table 47-1 shows that for $n = 1$ (the ground state) this probability is 2%. For $n = 2$ and $n = 3$, it is 10% and 30%, respectively.

We can understand this variation if we realize that the closer the energy E of the electron is to U_0, the closer the electron is to being a free particle, for which there is no localization and hence no quantization. Speaking very loosely indeed, we can perhaps imagine that an electron close to the top of the well is almost a free particle, so it "spreads out." The electron in the state with $n = 1$, however, is close to the bottom of the well and "thinks" that it is trapped in a well that is almost infinite in depth, so it senses rigid walls.

Third property: In Table 47-1 we see that the energies for an electron trapped in a finite well are significantly lower than those for an infinite well of the same width. For the ground state, for example, we have $E_1 = 37.6$ eV for the infinite well and $E_1 = 24$ eV for the finite well.

We can understand this by examining the probability density curves for the two ground or $n = 1$ states. For the infinite well (see Fig. 47-4*a*, $n = 1$) half a de Broglie wave fits exactly between the rigid walls of the well. For the finite well, however (see Fig. 47-7, $n = 1$) a half-wave is larger than the width of the well; there is "something left over." If the wavelength of the ground-state matter wave is larger for the finite well, it follows from the de Broglie

TABLE 47-1 **Energy Levels in a Finite and in an Infinite Well**

	Infinite Well[1]	*Finite Well*[2]	
Quantum Number n	*Energy* (eV)	*Energy* (eV)	*Probability of Being Outside the Well*
5	940	—	—
4	602	—	—
3	338	200	30%
2	150	94	10%
1	37.6	24	2%

[1] See Figure 47-3 and Sample Problem 47-1.

[2] See Figure 47-6.

relation (Eq. 46-1, $\lambda = h/p$) that it must have a smaller momentum and thus a smaller energy, just as we observe.

A Quantum Corral

When a scanning tunneling microscope (see Section 46-7) is in operation, its tip exerts a small force on isolated atoms that may be located on an otherwise smooth surface. By careful manipulation of the position of the tip, such isolated atoms can be "dragged" across the surface and deposited at another location. Using this technique, scientists at IBM's Almaden Research Center moved individual iron atoms across a carefully prepared copper surface, forming the atoms into a circle, which they named a *quantum corral;* see Fig. 47-8. Each iron atom in the circle is nestled in a hollow in the copper surface, equidistant from three nearest-neighbor copper atoms. The corral was assembled at a low temperature (about 4 K) to minimize the tendency of the iron atoms to move randomly about on the surface because of their thermal energies.

The ripples within the corral are due to matter waves associated with electrons that can move over the copper surface but are largely trapped in the potential well formed by the corral. The two-dimensional wave functions and probability densities can be calculated using methods similar to those we have used in this section for the one-dimensional finite well. The pattern of the ripples is in excellent agreement with the predictions of quantum mechanics for the electron probability densities.

47-4 AN ELECTRON TRAPPED IN AN ATOM

Now we are ready to study an electron trapped in an atom. For simplicity, we choose the simplest atom—hydrogen—which contains only one electron. To define the problem we must identify the potential energy function for this electron, just as we did in Fig. 47-6*a* for a finite well. In that case the motion of the electron was one-dimensional. The electron in the hydrogen atom, however, can move in three dimensions. Its potential energy, conveniently, involves only a single variable—namely, the radial distance r from the nucleus to the electron.

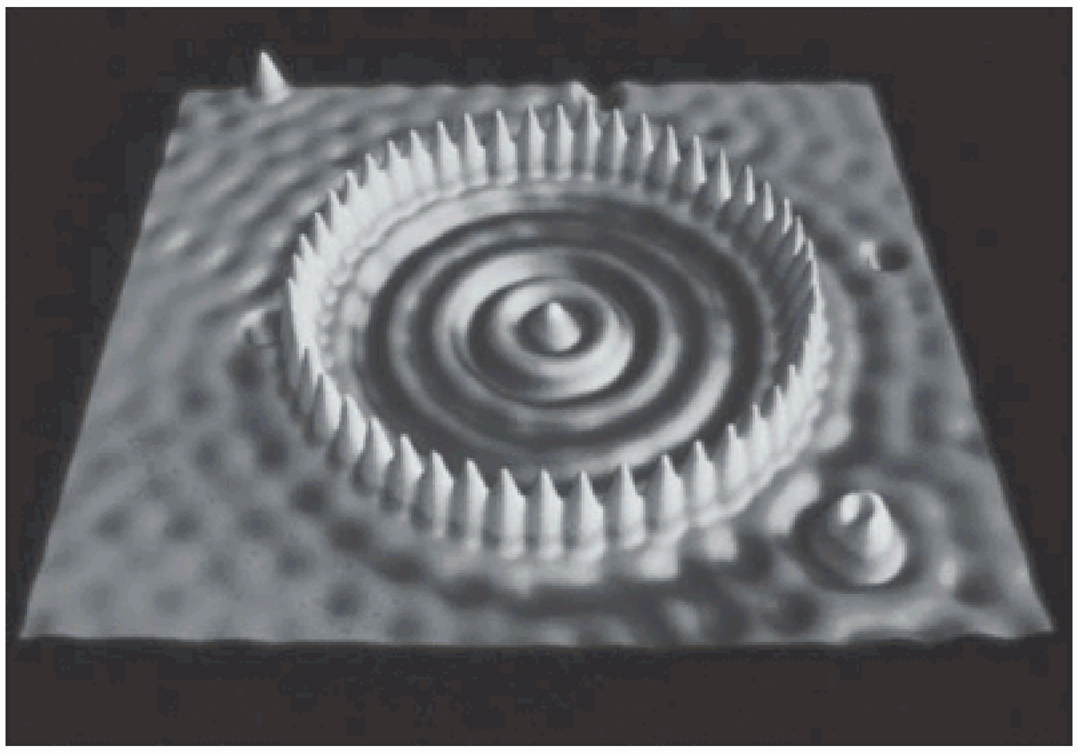


FIGURE 47-8. A quantum corral. Visit www.almaden.ibm.com/vis/index.html for a gallery of corrals and other STM displays.

The potential energy function follows from Coulomb's law and is

$$U(r) = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r}, \qquad (47\text{-}18)$$

in which $+e$ is the charge of the central proton and $-e$ is that of the electron. Figure 47-9 shows a plot of $U(r)$; it too, is a potential well. In the figure we have expressed the radial separation as a dimensionless ratio r/a_0. The constant a_0, called the *Bohr radius,* is a convenient unit of length for distances on the atomic scale. Its numerical value, as we discuss later in this section, is 52.9 pm, which is approximately the radius of a typical atom.

In the next section we will begin to consider the solutions to Schrödinger's equation for the hydrogen atom. First, however, we give a brief review of the historical background.

The Historical Background

In the first quarter of the 20th century one of the major unsolved problems in physics was, "What is the structure of an atom?" There was much experimentally obtained information demanding to be interpreted. In particular, atoms were known to emit (and absorb) light only at certain sharply defined wavelengths. Figure 47-10*a*, for example,

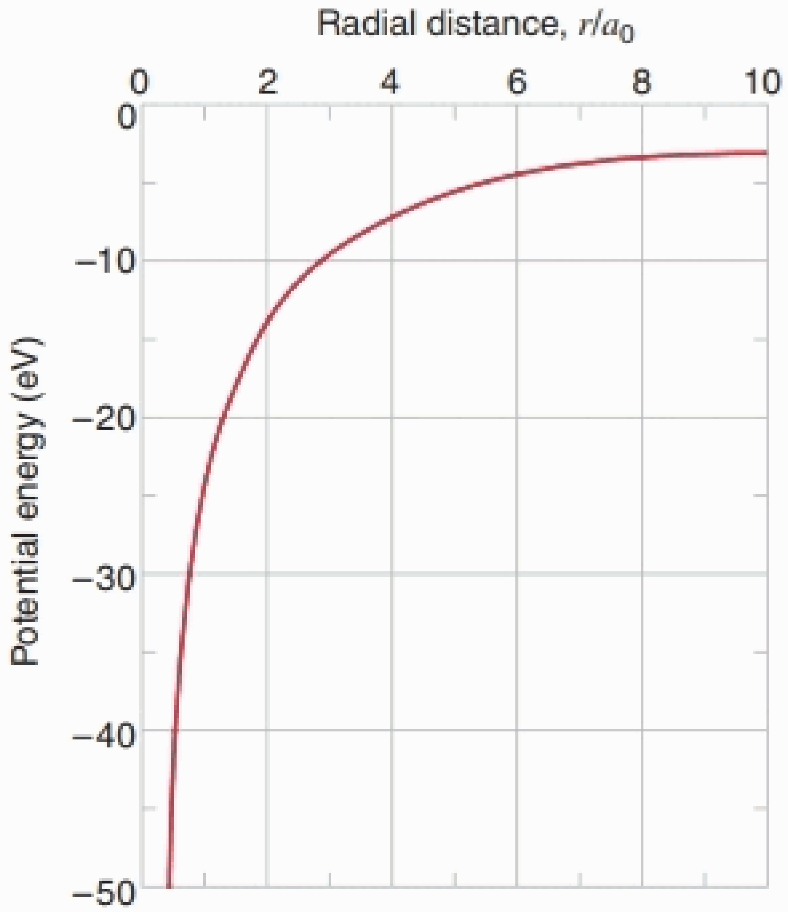


FIGURE 47-9. The potential energy function for the electron in a hydrogen atom; see Eq. 47-18. The potential energy $U(r)$ is plotted against the dimensionless ratio r/a_0, where a_0 is a convenient standard length called the *Bohr radius.*

shows a portion—called the Balmer series—of the so-called *line spectrum* of the hydrogen atom, which might be observed by using a diffraction grating to display the wavelengths of the light emitted by hydrogen gas. Figure 47-10*b* shows schematically that the Balmer series is only part of a more extensive spectrum, made up of several discrete series of spectrum lines, each series named after an early investigator. Note that each series occupies a finite section of the electromagnetic spectrum. For the Balmer series of Fig. 47-10*a*, the line with the longest wavelength is marked H_α and has the value $\lambda = 656.28$ nm. The short-wavelength series limit of the Balmer series, marked with an arrow in Fig. 47-10*a*, has the value $\lambda = 364.70$ nm.

Physicists, working with the frequencies (rather than the wavelengths) of the spectrum lines, tried to relate these frequencies to the frequency with which the electron circulates in an orbit around its nucleus. Unfortunately, this "solar system" model of an atom cannot explain the spectral properties of hydrogen without some additional assumptions. Let us see why.

An accelerating charge radiates energy into the surrounding space. A familiar example is the energy emitted by a radio transmitting antenna, associated with the charges that surge back and forth with accelerated motion within the body of the antenna. An electron, presumed to move in a circular orbit around an atomic nucleus, is accelerated, as are all bodies that move in circular arcs. The electron, because of this acceleration, will radiate away its energy as it moves. Such orbiting electrons, as they radiate away their energy, would rapidly spiral down into the nucleus. An atom based on classical orbiting electrons would not emit sharply defined spectrum lines and its orbit would not be stable.

Bohr's Model of the Hydrogen Atom

In 1913, more than a decade before quantum mechanics was established, the Danish physicist Niels Bohr (1885–1962) proposed a model of the hydrogen atom based on a clever hybrid combination of classical and early quantum physics. Although Bohr's model was later replaced by a fully developed quantum model, it nevertheless served as a powerful stimulus of later developments.

Even though Bohr knew that, according to classical physics, electron orbits are not stable, he asked himself a bold question. What if classical physics is wrong and the orbits actually *are* stable? In modern language, he made this bold postulate:

> *Electrons in an atom can exist in stable states of fixed energy, without radiating.*

Thus, at a stroke, Bohr advanced the concept of a *quantum state,* a concept that we have made much use of in our examination of electrons trapped in wells.

Bohr then made a second, equally bold, postulate; namely,

> *An atom emits (or absorbs) radiation at a fixed frequency when an electron in that atom moves from one of its stable quantum states to another.*

The frequency f of the emitted (or absorbed) radiation is given by the *Bohr frequency condition,* which is

$$E_{n_i} - E_{n_f} = hf, \tag{47-19}$$

in which h is the Planck constant, and E_{n_i} is the higher and E_{n_f} the lower of the energies of the two quantum states (n_i and n_f) involved in the transition.

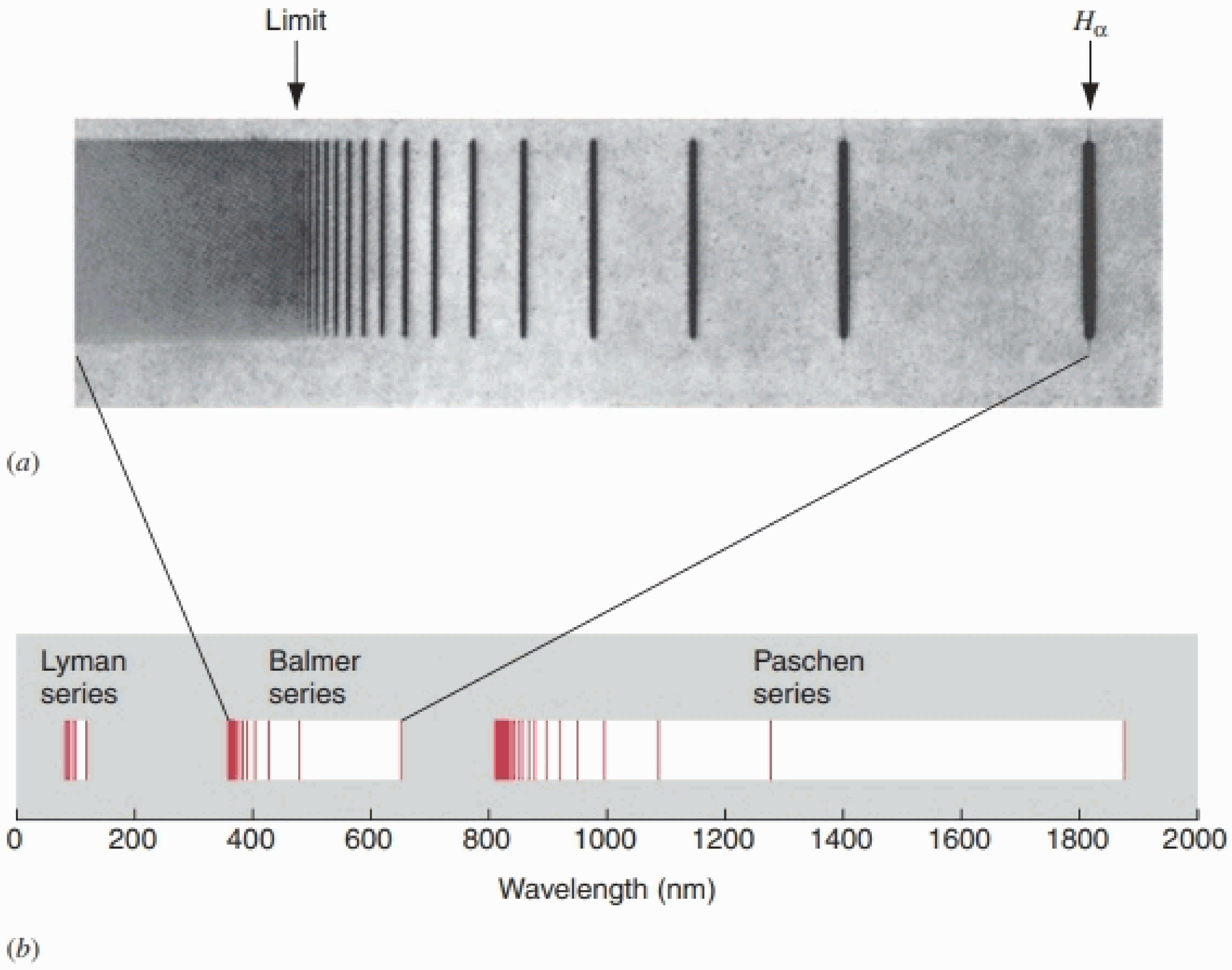


FIGURE 47-10. (*a*) A series of spectrum lines, called the *Balmer series,* emitted by the hydrogen atom in the visible and the near ultraviolet region of the electromagnetic spectrum. (*b*) A more complete spectrum of the hydrogen atom, including the Lyman series and the Paschen series. Two other series, the Brackett and the Pfund series, exist in the far infrared at wavelengths beyond the scale of the figure.

With these assumptions, Bohr reduced the problem of the hydrogen atom to that of finding the energies of its quantum states. With these in hand, the frequencies—and thus also the wavelengths—of the spectrum lines can be calculated, using Eq. 47-19.

To make quantitative predictions, Bohr advanced still another useful concept, the *correspondence principle,* which we first discussed in Section 47-2. Its essence is that the quantum view must merge smoothly into the classical view as the quantum numbers increase. By making clever use of this principle, Bohr was able to create a model in which the quantum states were represented by a discrete set of circular orbits, whose radii were given by

$$r_n = \left(\frac{\epsilon_0 h^2}{\pi m e^2}\right) n^2 = a_0 n^2 \qquad n = 1, 2, 3, \ldots, \qquad (47\text{-}20)$$

in which n is a quantum number, m is the electron mass, and ϵ_0 is the electric constant. That is, an electron is permitted to be in an orbit with radius a_0, $4a_0$, $9a_0$, . . . , but never with radius $2a_0$ or $6.1a_0$.

Equation 47-20 introduces a_0, the Bohr radius that we mentioned above in connection with Fig. 47-9. We see that it is the radius of the smallest allowed orbit. Its numerical value may be calculated from Eq. 47-20. Using four significant figures for the constants involved (see Appendix B) this value is

$$a_0 = \frac{\epsilon_0 h^2}{\pi m e^2} = \frac{(8.854 \times 10^{-12}\ \text{F/m})(6.626 \times 10^{-34}\ \text{J}\cdot\text{s})^2}{(\pi)(9.109 \times 10^{-31}\ \text{kg})(1.602 \times 10^{-19}\ \text{C})^2}$$
$$= 5.29 \times 10^{-11}\ \text{m} = 52.9\ \text{pm}.$$

Bohr also calculated that the energies of the orbital quantum states are given by

$$E_n = -\frac{me^4}{8\epsilon_0^2 h^2}\frac{1}{n^2} \qquad n = 1, 2, 3, \ldots. \qquad (47\text{-}21)$$

The energies are negative because the configuration of zero potential energy has been defined as that in which the electron is at a very large distance from the nucleus and is at rest. Equation 47-21 does indeed verify that $E_n \to 0$ as $n \to \infty$. Bohr and others made great use of this atom model, using it to provide other information about the hydrogen atom, such as the angular momenta of the allowed orbital states, and extending it usefully to atoms beyond hydrogen in the periodic table.

When Bohr established his model for the hydrogen atom, Schrödinger's equation had not yet been advanced and the uncertainty principle was some years in the future. All he had to work with was Planck's quantization concept and Einstein's photon concept. Under these circumstances, it is perhaps not surprising that some of his conclusions did not endure. In particular, we no longer imagine that the electrons in an atom travel in well-defined orbits. The very idea of "orbit" violates Heisenberg's uncertainty principle because it conjures up the image of a particle of fixed momentum occupying a known position at any given instant; these two quantities (position and momentum) cannot be known simultaneously with unlimited precision. Unfortunately, the popular image of an atom continues to propagate this incorrect orbital model.

However, much of Bohr's predictions did endure. His equation for the energies of the hydrogen atom states (Eq. 47-21) is the same as that predicted from a solution of Schrödinger's equation. Furthermore, the very concept of the existence of quantum states of fixed energy and the validity of the Bohr frequency condition remain with us as part of the firm foundation of modern quantum physics. Bohr's theory provided an extremely useful bridge connecting classical and quantum ideas and served as a stimulus to the full development of quantum mechanics as we know it today.

Figure 47-11 shows the energy levels of the hydrogen atom, calculated using Eq. 47-21. The vertical lines, drawn using Eq. 47-19, show the energies of the emitted photons corresponding to the spectrum lines shown in Fig. 47-10. Note that each series of spectrum lines corresponds to different choices (in Eq. 47-19) of the initial state n_i for the same final state n_f. For example, the spectrum lines in which the electron ends up in the quantum state with $n_f = 2$ form the Balmer series. The vertical arrow in the Balmer series, marked H_α in Figs. 47-10*a* and 47-11, corresponds to the spectrum line in that series with the smallest frequency or, correspondingly, the largest wavelength. The series limits for the Balmer series corresponds to transition from a state with $n_i = \infty$ to $n_f = 2$.

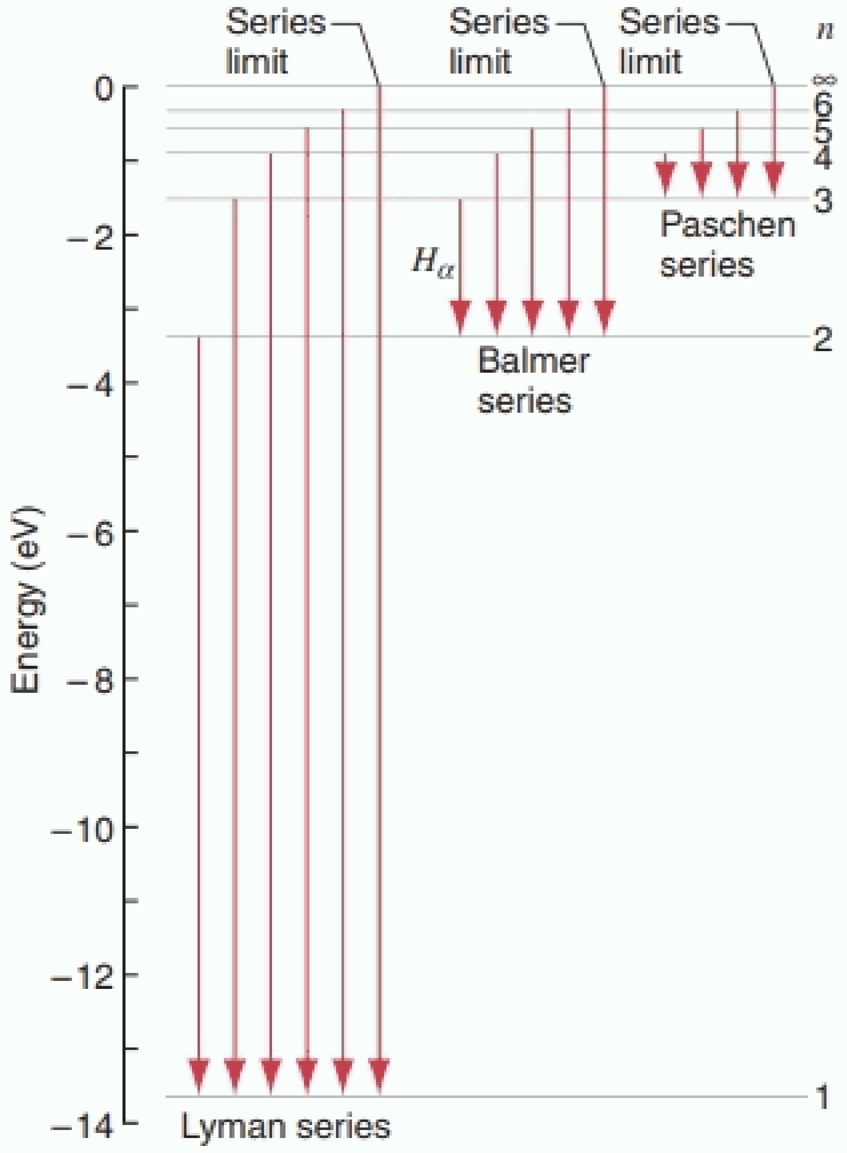


FIGURE 47-11. The horizontal lines show the energy levels of the hydrogen atom. The vertical arrows, grouped into series, show the transitions that occur between these levels. Arrows pointing *downward* correspond to the *emission* of energy by the atom.

SAMPLE PROBLEM 47-5. Calculate the wavelength of the spectrum line marked H_α in Fig. 47-10*a*.

Solution This line is in the Balmer series and, from Fig. 47-11, we see that it arises when the electron in the hydrogen atom moves from the quantum state with $n_\text{i} = 3$ to that with $n_\text{f} = 2$. By combining Eqs. 47-19 and 47-21 we can find the frequency of the Balmer line and, knowing that, we can easily find its wavelength.

Because we wish to find the wavelength to four significant figures, we must use values of the fundamental constants such as m and h that are precise to five significant figures; see Appendix B. From Eq. 47-19 we have

$$hf = E_3 - E_2$$

or

$$f = \frac{E_3 - E_2}{h}.$$

From Eq. 47-21 we then have

$$\begin{aligned} f &= -\frac{me^4}{8\epsilon_0^2 h^3}\left(\frac{1}{3^2} - \frac{1}{2^2}\right) \\ &= -\frac{(9.1094 \times 10^{-31}\ \text{kg})(1.6022 \times 10^{-19}\ \text{C})^4}{(8)(8.8542 \times 10^{-12}\ \text{F/m})^2(6.6261 \times 10^{-34}\ \text{J}\cdot\text{s})^3}(-0.13889) \\ &= 4.5695 \times 10^{14}\ \text{Hz}. \end{aligned}$$

The corresponding wavelength of this line is given by

$$\begin{aligned} \lambda = \frac{c}{f} &= \frac{2.9979 \times 10^8\ \text{m/s}}{4.5695 \times 10^{14}\ \text{Hz}} \\ &= 6.561 \times 10^{-7}\ \text{m} = 656.1\ \text{nm}. \end{aligned}$$

This value, calculated from Bohr's theory, agrees well with experiment.

SAMPLE PROBLEM 47-6. A neutral helium atom has two electrons. Suppose one of them is removed, leaving the helium ion He^+. This ion, with its single electron, resembles the hydrogen atom, except that its nuclear charge is $+2e$ rather than $+e$. (*a*) Find the radius of the electron orbit of the ground state of the helium ion according to the Bohr model. (*b*) Find the wavelength of the photon emitted when the electron makes a transition from the $n = 3$ orbit to the $n = 2$ orbit.

Solution (*a*) Equation 47-20 for the orbital radii is valid only in the case of a nuclear charge of $+e$. Let us first obtain a more general expression when the nuclear charge is $+Ze$. Equation 47-20 includes the factor e^2. This factor originates from the expression for the Coulomb force between the electron and the nucleus in hydrogen: $F = e^2/4\pi\epsilon_0 r^2$. In the case of an atom with a single electron and a nuclear charge of Ze, the force would be $F = Ze^2/4\pi\epsilon_0 r^2$, because the product of the two charges is now $(e)(Ze)$. That is, the force must be modified by replacing the factor e^2 by the factor Ze^2. Making the same change in Eq. 47-20, we obtain a general expression for the radii of the Bohr orbits in any single-electron ion with a nuclear charge of Ze:

$$r_n = \left[\frac{\epsilon_0 h^2}{\pi m (Ze^2)}\right] n^2 = \frac{a_0 n^2}{Z} \qquad n = 1, 2, 3, \ldots . \qquad (47\text{-}22)$$

Comparison with Eq. 47-20 shows that the orbits are smaller than the corresponding orbits in hydrogen by the factor $1/Z$.

For the ground state ($n = 1$) of the helium ion ($Z = 2$), we have

$$r_2 = \frac{(52.9\ \text{pm})(1^2)}{2} = 26.5\ \text{pm}.$$

(*b*) The energy levels in hydrogen according to the Bohr model are given in Eq. 47-21. If we make the same replacement of e^2 by Ze^2, we must replace the factor $e^4 = (e^2)^2$ in Eq. 47-21 by $(Ze^2)^2 = Z^2e^4$. Making this substitution in Eq. 47-21, we obtain a general expression for the energy levels in any single-electron ion with nuclear charge Ze:

$$E_n = -\frac{mZ^2e^4}{8\epsilon_0^2 h^2}\frac{1}{n^2} \qquad n = 1, 2, 3, \ldots . \qquad (47\text{-}23)$$

The calculation of the frequency of the transition from $n = 3$ to $n = 2$ in helium can be carried out using exactly the same method we used for hydrogen in Sample Problem 47-5. The only difference will be the additional factor of $Z^2 = 4$ that will appear in the expressions for the helium ion. That is, $f_\text{He} = 4f_\text{H} = 4(4.5695 \times 10^{14}\ \text{Hz}) = 1.8278 \times 10^{15}$ Hz. The corresponding wavelength is

$$\lambda = \frac{c}{f} = \frac{2.9979 \times 10^8\ \text{m/s}}{1.8278 \times 10^{15}\ \text{Hz}} = 1.640 \times 10^{-7}\ \text{m} = 164.0\ \text{nm}.$$

Compared with the same transition in hydrogen, the frequency is larger by the factor Z^2 and the wavelength is smaller by the factor $1/Z^2$.

47-5 THE GROUND STATE OF THE HYDROGEN ATOM

We now discuss the solution of the Schrödinger equation corresponding to the potential energy, given in Eq. 47-18, that represents the Coulomb interaction between the electron and the nucleus. We shall not give the mathematical details of solving this three-dimensional equation, which are beyond the level of this text. Instead, we will merely present the solutions for the wave functions and discuss their meaning.

Because the potential energy for the hydrogen atom depends only on r (the distance of the electron from the central proton), the problem is most easily expressed in a spherical polar coordinate system in which the proton is fixed at the origin. Then the electron wave functions can be written in terms of the coordinates r, θ, ϕ instead of x, y, z. In one dimension, the probability density $\psi^2(x)$ gave the probability *per unit length* for the electron to be found in a small element of length at the coordinate x. In three dimensions, the probability density $\psi^2(r, \theta, \phi)$ gives the probability *per unit volume* for the electron to be found in a small volume element at the coordinates r, θ, ϕ. In other words, $\psi^2(r, \theta, \phi)\, dV$ gives the probability of finding the electron in the volume element dV.

The wave function for the ground state turns out to depend on only the single variable r. The expression is

$$\psi(r) = \frac{1}{\sqrt{\pi a_0^3}}\, e^{-r/a_0}, \qquad (47\text{-}24)$$

in which a_0 is the Bohr radius. This wave function is spherically symmetrical; the variables θ and ϕ are not involved.

The probability density for the ground state is

$$\psi^2(r) = \frac{1}{\pi a_0^3} e^{-2r/a_0}. \tag{47-25}$$

Figure 47-12*a* is a "dot plot" representation of Eq. 47-25, the density of the dots suggesting the probabilistic nature of the electron's location. It is fundamental to quantum mechanics that the question, "Where is the electron at this instant?" has no meaning and must be replaced by, "What is the probability that, at this instant and at this location, we can detect the presence of an electron?" Note that, in the quantum picture, there is no mention of orbits for the electrons.

A more useful quantity for representing the electron's position is the *radial probability density* $P(r)$. This is defined so that $P(r)\,dr$ gives the probability that, regardless of direction, the electron will be found to lie between two spheres whose radii are r and $r + dr$. The volume dV between these spheres is $(4\pi r^2)(dr)$, so that we can write

$$P(r)\,dr = \psi^2(r)\,dV = \psi^2(r)(4\pi r^2)(dr).$$

Substituting from Eq. 47-25 then leads to

$$P(r) = \psi^2(r)(4\pi r^2) = \frac{4}{a_0^3} r^2 e^{-2r/a_0}. \tag{47-26}$$

Figure 47-12*b* shows a plot of $P(r)$. In Sample Problem 47-7 we show that the maximum value of $P(r)$ occurs at $r = a_0$. Equation 47-26 provides the basis for the modern interpretation of the Bohr radius a_0. It is the radius at which you are most likely to find the electron when the hydrogen atom is in its ground state.

Figure 47-12*b* suggests that the question, "How large is the hydrogen atom?" has no firm answer. However, the so-called 90% radius, indicated in that figure, gives a rough measure of atomic size. It is the radius of a sphere such that there is a 90% probability that the hydrogen atom's electron will be found inside the sphere and, of course, a 10% chance that it will be found outside. The 90% radius turns out to be 2.7 Bohr radii or 1.4×10^{-10} m = 140 pm.

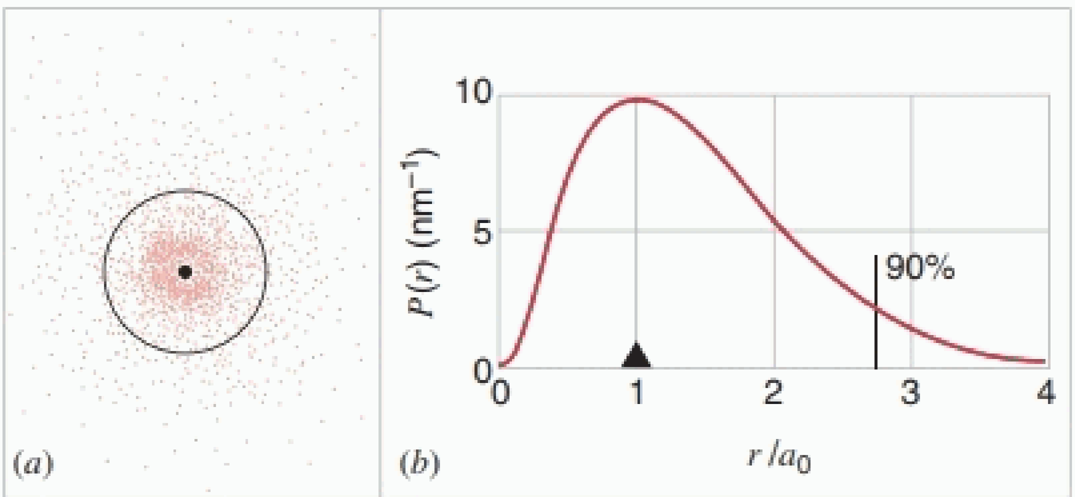


FIGURE 47-12. (*a*) A "dot plot" representation of the electron probability density for the ground state of the hydrogen atom; see Eq. 47-25. A circle has been drawn at radius $r = a_0$. (*b*) The radial probability density, given by Eq. 47-26. The small filled triangle shows that the maximum probability occurs at $r = a_0$.

SAMPLE PROBLEM 47-7. Verify that the maximum of the radial probability density curve of Fig. 47-12*b* falls at $r = a_0$, where a_0 is the Bohr radius.

Solution Differentiating Eq. 47-26 with respect to r leads to

$$\frac{dP}{dr} = \frac{4}{a_0^3}(2r)\,e^{-2r/a_0} + \frac{4}{a_0^3} r^2 \left(-\frac{2}{a_0}\right) e^{-2r/a_0}$$
$$= \frac{8r}{a_0^3} e^{-2r/a_0}\left(1 - \frac{r}{a_0}\right).$$

At the maximum of the curve we must have $dP/dr = 0$ and, by simple inspection, we see that this indeed occurs for $r = a_0$. Note that we also have $dP/dr = 0$ at $r = 0$ and as $r \to \infty$. As Fig. 47-12*b* shows, however, these do not correspond to maxima in the curve. You can verify this by finding the second derivative d^2P/dr^2, which must be negative for a maximum.

47-6 ANGULAR MOMENTUM OF ELECTRONS IN ATOMS

Now that we have presented the ground-state wave function of the electron trapped in a hydrogen atom, we would like to make a similar discussion of the wave functions and probability densities for the excited states. However, when we solve the Schrödinger equation for the hydrogen atom, a three-dimensional problem, a new feature emerges that was not present when we considered solutions to the Schrödinger equation in one dimension. This new feature is related to the *angular momentum* of the electron.

For example, the solutions for the hydrogen atom in the first excited state ($n = 2$) show several different types of wave functions and corresponding probability densities, all of which have the same energy (E_2 in Eq. 47-21). One of these wave functions (let us call it state 2*A*) turns out to be spherically symmetric, just like the ground state (although its functional form is different from that of the ground-state wave function). For state 2*A*, the wave function ψ and the probability density ψ^2 depend only on r. There is another very different wave function with energy E_2; in this state (which we call for now state 2*B*), the wave function turns out to depend not only on r but also on the angular coordinates θ and ϕ, and the probability density ψ^2 depends on r and θ. (The probability density does not depend on ϕ and is thus cylindrically symmetric.*) In fact, the state 2*B* has two substates that represent different probability densities; in one of these substates, the electron is found with the greatest probability close to the *xy* plane (that is, with the polar angle θ near 90°), and in the other substate the electron is most likely found near the positive or negative *z* axis (θ near 0° or 180°). In the next section we will discuss these wave functions in more detail. In this section we discuss how the angular momentum helps us to label these different states and understand their properties.

---

* In actuality, the wave function ψ is complex, and the probability density should be written as $|\psi|^2$. However, the r and θ parts of the wave function for the hydrogen atom are real, and the ϕ part disappears when we calculate $|\psi|^2$. For simplicity, we will therefore write the probability densities for the hydrogen atom as ψ^2 and ignore the ϕ dependence of the wave function.

The Orbital Angular Momentum

We introduced angular momentum in Chapter 10 and discussed its importance in solving rotational problems. Classically, the angular momentum of a particle relative to a chosen origin O is represented as $\vec{\mathbf{L}} = \vec{\mathbf{r}} \times \vec{\mathbf{p}}$, where $\vec{\mathbf{r}}$ is the particle's displacement from the origin and $\vec{\mathbf{p}}$ is its linear momentum. As we have seen, the classical variables r and p cannot have their same meaning in quantum mechanics, and so it should not be surprising that angular momentum takes on a new meaning in quantum mechanics. Here we are referring to the angular momentum that an electron has because of its motion about the nucleus; we call this its *orbital angular momentum.* Later we will learn that the electron has yet another kind of angular momentum that has no classical counterpart.

When we solve the Schrödinger equation in three dimensions for the hydrogen atom, three quantum numbers emerge from the mathematical solution. One quantum number is the *principal quantum number n,* which is identical with the quantum number n associated with labeling the states in the Bohr model of the hydrogen atom. The allowed values of the principal quantum number are $n = 1, 2, 3, \ldots$, just as in the Bohr model. The energy levels calculated from solving the Schrödinger equation are identical with those of the Bohr model (Eq. 47-21). However, as we have seen already in the case of the ground state, the Bohr formula for the radii (Eq. 47-20) has no meaning for the quantum-mechanical solution, because the radial probability density has nonzero values that span the full range from $r = 0$ to $r = \infty$. In the quantum-mechanical picture, we should instead be speaking of the average or most probable radius; this quantity does increase with increasing values of n, although not at all in the simple way that the Bohr model would predict.

The second quantum number that emerges from the solution to the Schrödinger equation is the *angular momentum quantum number l.* The mathematical solution shows that l is permitted to take integer values as follows:

$$l = 0, 1, 2, \ldots, n - 1. \tag{47-27}$$

That is, for each n, the values of l range from 0 to $l_{\max} = n - 1$. For example, for the ground state ($n = 1$), $l_{\max} = 0$. Thus the only allowed value of l is 0. For the first excited state ($n = 2$), $l_{\max} = 1$, and so the permitted values are $l = 0$ and $l = 1$. We will see in the next section that these two values of l correspond exactly to the two states we called 2*A* and 2*B* in our discussion earlier in this section.

It is often convenient to label the states by giving the principal quantum number n and angular momentum quantum number l. The value of l is given according to the following code*:

l	0	1	2	3	4	5
Code	*s*	*p*	*d*	*f*	*g*	*h*

Using this system, the ground state would be labeled 1*s* (for $n = 1$, $l = 0$). The first excited state has two designations: 2*s* ($n = 2$, $l = 0$) and 2*p* ($n = 2$, $l = 1$). The latter two states are the ones we called respectively 2*A* and 2*B*.

The quantum mechanical calculation also gives us a value for the magnitude or length L of the angular momentum vector:

$$L = \sqrt{l(l + 1)}\,\frac{h}{2\pi}. \tag{47-28}$$

The Direction of $\vec{\mathbf{L}}$

The third quantum number that emerges from the solution of the Schrödinger equation is called the *magnetic quantum number m_l*. It can take values from $-l$ to $+l$ in integer steps, or

$$m_l = -l, -(l - 1), \ldots, -1, 0, +1, \ldots, +(l - 1), +l \tag{47-29}$$

Note that there are $2l + 1$ different possible values of m_l for each value of l. For example, for $l = 3$ there are seven allowed values of m_l; they are $-3, -2, -1, 0, +1, +2$, and $+3$.

The magnetic quantum number specifies the z component of the angular momentum vector $\vec{\mathbf{L}}$:

$$L_z = m_l\,\frac{h}{2\pi}. \tag{47-30}$$

If we know the magnitude of the vector $\vec{\mathbf{L}}$ from Eq. 47-28 and its z component from Eq. 47-30, we can determine the angle θ between $\vec{\mathbf{L}}$ and the z axis:

$$\theta = \cos^{-1}\frac{L_z}{L} = \cos^{-1}\frac{m_l}{\sqrt{l(l + 1)}}. \tag{47-31}$$

Equation 47-31 shows that not all angles θ are permitted to exist—only those corresponding to the allowed values of m_l. This restriction on the directions of the angular momentum vector $\vec{\mathbf{L}}$ is called *space quantization.* Space quantization does not exist in classical physics; the angular momentum of a classical orbiting particle can have any direction in space. Only in quantum mechanics does this restriction

---

* These abbreviations date from the early days of atomic spectroscopy before quantum physics. The letters *s, p, d, f* stand for sharp, principal, diffuse, and fundamental, which were terms used by spectroscopists to describe atomic spectra. After *f*, the indices are in alphabetical order

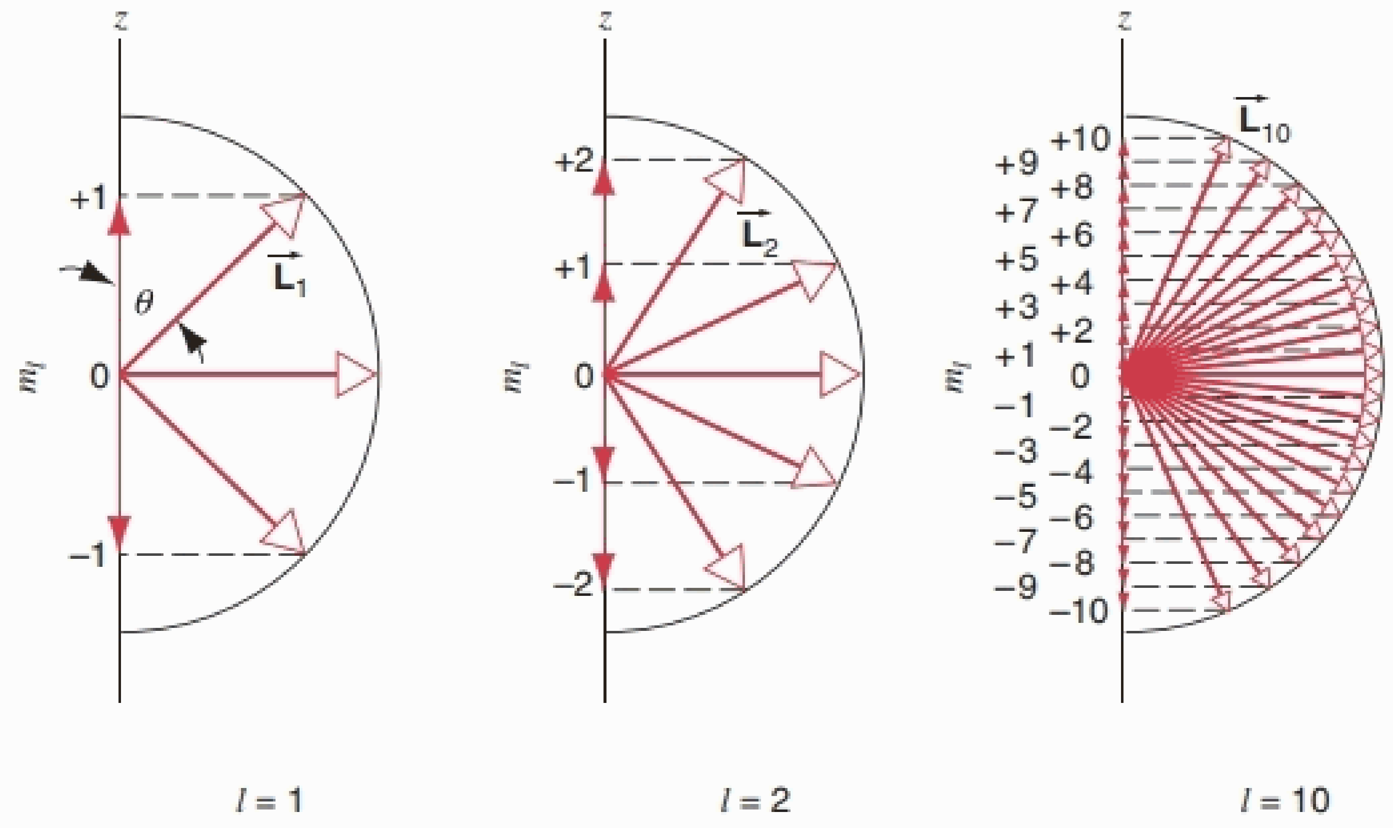


FIGURE 47-13. The allowed projections of the angular momentum vector $\vec{L}$ on the z axis for $l = 1, 2,$ and 10. The numbers on this axis are the corresponding value of the magnetic quantum number m_l. The three figures are drawn to different scales.

arise. Figure 47-13 shows the allowed values of L_z and θ for $l = 1, 2,$ and 10.

Because the Coulomb potential in the hydrogen atom is spherically symmetric, we are free to set up our coordinate system in any orientation we choose. The direction of the z axis is then completely arbitary. Equation 47-31 applies to *any* choice of the z axis. To observe the effect of spatial quantization, we must break the spherical symmetry of the atom—for example, by placing it in a magnetic field whose direction defines the z axis (hence the name "magnetic" quantum number for m_l). In the absence of any such preferred direction, the different m_l values do not correspond to different states of motion of the electron and thus do not affect the energy of the atom or the probability density of the electron.

Angular Momentum and the Uncertainty Principle

$\vec{L}$ cannot be fully aligned with the z axis, the smallest value of θ occurring when $m_l = l$ in Eq. 47-31. This fact is consistent with Heisenberg's uncertainty principle. In Eq. 46-8, we wrote this principle, as it applies to the z component of the motion of a free particle, as

$$\Delta p_z \cdot \Delta z > h/2\pi. \tag{47-32}$$

By analogy, we can recast this equation in angular form as

$$\Delta L_z \cdot \Delta \phi > h/2\pi. \tag{47-33}$$

Here we have replaced Δp_z, the uncertainty of the component of *linear* momentum in the z direction, with ΔL_z, the uncertainty in the component of *angular* momentum in the z direction. We have also replaced Δz, the uncertainty in *linear* position in the z direction, with $\Delta \phi$, the uncertainty in *angular* position around the z axis.

Equation 47-30 tells us that L_z is precisely known once we have specified the magnetic quantum number m_l. It follows that ΔL_z, the uncertainty in L_z, must be zero. Equation 47-33 then requires that $\Delta \phi \rightarrow \infty$, which means that we have no information at all about the angular position of $\vec{L}$ as it precesses about the z axis. This, in turn, means that we have no information at all about the components L_x and L_y of the angular momentum vector. We can know the magnitude of $\vec{L}$ and its projection L_z in the z direction, but quantum mechanics does not permit us to have any additional knowledge of the direction of $\vec{L}$.

SAMPLE PROBLEM 47-8. Find the minimum value of θ in Fig. 47-13 for $l = 1, 10^2, 10^3, 10^4,$ and 10^9.

Solution Convince yourself that the minimum value occurs when we put $m_l = l$ in Eq. 47-31. Doing so and rearranging leads to

$$\theta_{\min} = \cos^{-1} \frac{l}{\sqrt{l(l+1)}} = \cos^{-1} \left[1 + \frac{1}{l}\right]^{-1/2}.$$

We see by inspection that if we let $l \rightarrow \infty$ then $\theta \rightarrow \cos^{-1} 1 = 0$. This is just what we expect from the correspondence principle, which asserts that, at high enough quantum numbers, quantum mechanics agrees with classical mechanics, which permits $\theta = 0$.

Substituting for l in the above equation leads to

l	$\theta_{\min}$
1	45.0°
10^2	5.7°
10^3	1.8°
10^4	0.57°
10^9	0.0018°

For a spinning macroscopic object, a compact disk for example, l would be enormously larger than 10^9, and $\theta_{\min}$ would be so close to zero that the difference would be far beyond the possibility of measurement. Thus as the angular momentum of a spinning object gets larger and larger, the space quantization of quantum mechanics merges gently into the continuous distribution of classical mechanics. We see once again how the correspondence principle operates.

Computational note: If you use the above formula to calculate $\theta_{\min}$ for $l = 10^9$, your calculator will probably overflow. Take advantage of the fact that $(1/l) \ll 1$ and develop an approximate formula for $\cos \theta_{\min}$. You will need to use both the binomial expansion and the series expansion for $\cos \theta$; see Appendix I.

SAMPLE PROBLEM 47-9. (*a*) For $n = 4$, what is the largest allowed value of l? (*b*) What is the magnitude of the corresponding angular momentum? (*c*) How many different components on the z axis may this angular momentum vector have? (*d*) What is the magnitude of the largest z component?

Solution (*a*) From Eq. 47-27, the largest value of l is $n - 1$, so $l_{\max} = 3$.
(*b*) From Eq. 47-28 we have

$$L = \sqrt{l(l+1)}\,\frac{h}{2\pi} = \sqrt{3(3+1)}\,\frac{6.63 \times 10^{-34}\ \mathrm{J\cdot s}}{2\pi}$$

$$= 3.66 \times 10^{-34}\ \mathrm{J\cdot s}.$$

Convince yourself that the Planck constant has the same dimensions as angular momentum.
(*c*) The number of components that the angular momentum vector may have on the z axis is equal to the number of the allowed values of the magnetic quantum number m_l. From Eq. 47-29 this number is $2l + 1$ or $2 \times 3 + 1 = 7$.
(*d*) The largest projected component is found from Eq. 47-30, in which the magnetic quantum number m_l has its largest value. From Eq. 47-29 this largest value is simply l, so we have

$$L_z = l\left(\frac{h}{2\pi}\right) = \frac{3h}{2\pi} = \frac{3(6.63 \times 10^{-34}\ \mathrm{J\cdot s})}{2\pi} = 3.17 \times 10^{-34}\ \mathrm{J\cdot s}.$$

Note from part (*b*) that this is smaller than the magnitude of the angular momentum vector ($= 3.66 \times 10^{-34}$ J·s), as it must be.

47-7 AN EXCITED STATE OF THE HYDROGEN ATOM

The state just above the ground state in energy corresponds to $n = 2$ and is called the *first excited state*. Its energy ($= -3.40$ eV) is found by putting $n = 2$ in Eq. 47-21. As Eq. 47-27 reminds us, there are two different angular momentum quantum numbers permitted for $n = 2$, namely, $l = 0$ and $l = 1$. In our notation, these are, respectively, the $2s$ and the $2p$ states. We deal with each in turn.

The 2*s* State

The wave function for this state, found in the process of solving Schrödinger's equation, turns out to be

$$\psi_{200}(r) = \frac{1}{4\sqrt{2\pi a_0^3}}\,(2 - r/a_0)\, e^{-r/2a_0}, \qquad (47\text{-}34)$$

in which the three subscripts 200 represent the quantum number sequence $n = 2$, $l = 0$, and $m_l = 0$. This wave function involves the variable r only; like the ground state, this state has spherical symmetry.

The probability density $\psi^2(r)$ and the radial probability density $P(r)$ for the $2s$ state are given by

$$\psi^2(r) = \frac{1}{32\pi a_0^3}\,(2 - r/a_0)^2\, e^{-r/a_0} \qquad (47\text{-}35)$$

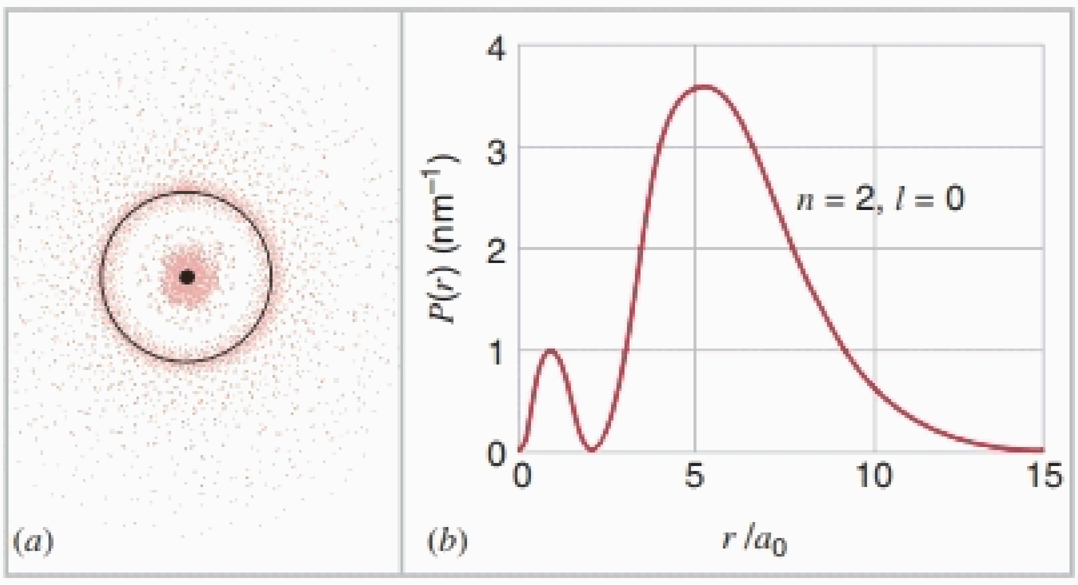


FIGURE 47-14. (*a*) A dot-plot representation of the electron probability density for a hydrogen atom in a $2s$ state; see Eq. 47-35. A circle has been drawn at radius $r = 4a_0$. (*b*) The radial probability density; see Eq. 47-36.

and

$$P(r) = [\psi^2(r)][4\pi r^2]$$

$$= \frac{1}{8a_0}\,(r/a_0)^2\,(2 - r/a_0)^2\, e^{-r/a_0}. \qquad (47\text{-}36)$$

Figure 47-14*a* is a "dot plot" of Eq. 47-35, the probability density. Figure 47-14*b* is a plot of Eq. 47-36, the radial probability density. Note that the latter has two maxima and goes to zero at $r = 2a_0$, as inspection of Eq. 47-36 verifies.

The 2*p* State

The $1s$ state, with $l = 0$, has no orbital angular momentum. The $2p$ state, however, with $l = 1$, does have orbital angular momentum; its z-axis projection (see Eq. 47-30) is $m_l(h/2\pi)$, where m_l, the magnetic quantum number (see Eq. 47-29), can take on the values -1, 0, or $+1$. The wave functions for these three states that comprise the $2p$ state are *not* spherically symmetric. They are functions not only of r but also of the polar angle θ, which is defined in Fig. 47-15*a*.

Figure 47-15*a* shows "dot plots" of the three probability densities for the $2p$ state; namely,

$$\psi^2_{21-1}(r,\theta), \quad \psi^2_{210}(r,\theta), \quad \text{and} \quad \psi^2_{21+1}(r,\theta).$$

The subscripts represent the quantum number sequence n, l, and m_l. The plots for $m_l = -1$ and for $m_l = +1$ are identical and all three plots have rotational symmetry about the z axis.

You are entitled to be a little puzzled about the lack of spherical symmetry shown in the dot plots of Fig. 47-15*a*. After all, the potential energy function that we inserted into the Schrödinger equation depends only on r. How does it happen that the probability densities depend also on the angle θ?

The answer to this puzzle comes when we realize that the three states in question have the same energy and, for the isolated hydrogen atom, there is no way to separate them experimentally. The hydrogen atom can be found with equal probability in each of the three states shown in Fig. 47-15*a*, so it makes sense to deal with a *weighted average*

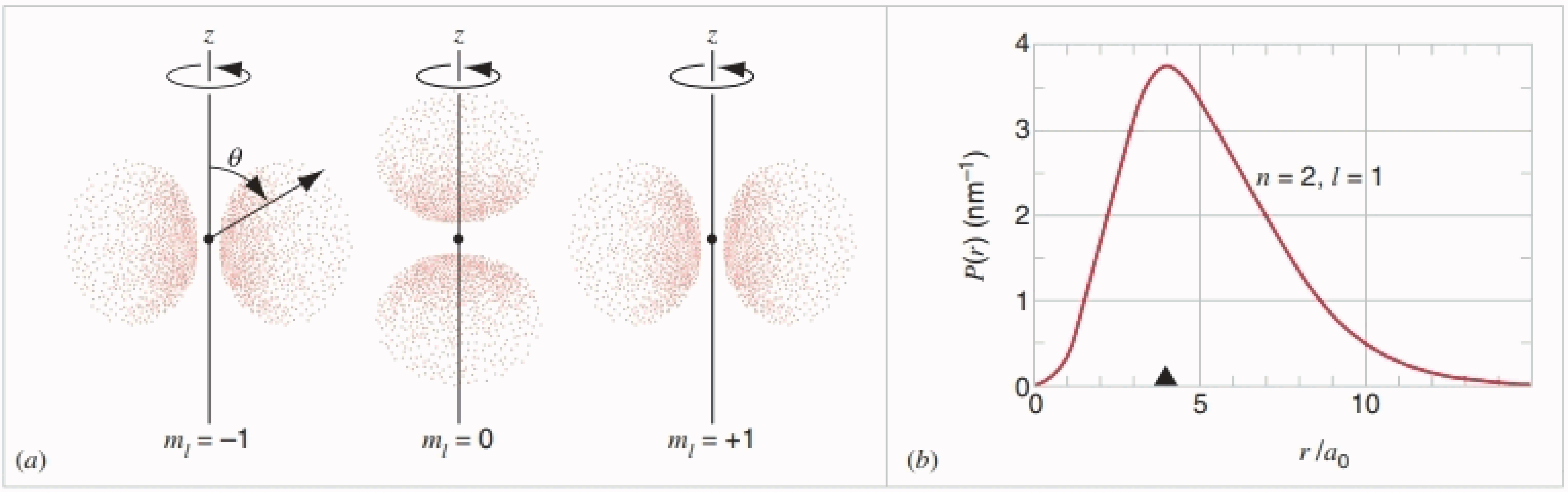


FIGURE 47-15. (*a*) Dot-plot representations of the electron probability density for a hydrogen atom in a $2p$ state. All three figures are symmetrical about the z axis. (*b*) The radial probability density for a weighted average of the three plots shown in (*a*). Note that the maximum probability occurs at $r = 4a_0$.

probability density for the $2p$ state as a whole. We can envision this graphically by considering that, in three dimensions, the dot plots corresponding to $m_l = -1$ and $m_l = +1$ in Fig. 47-15*a* are "doughnut-shaped." The dot plot corresponding to $m_l = 0$ has an upper and a lower lobe and it is reasonable that, when the three dot plots (each weighted by a factor of 1/3) are superimposed in three dimensions, something with spherical symmetry may well result.

That is indeed the case. The resulting probability density for the three superimposed plots turns out to be

$$\psi_{2p}^2(r) = \frac{1}{96\pi a_0^3}\, r^2\, e^{-r/a_0}. \qquad (47\text{-}37)$$

The subscript $2p$ on the probability density gives the value of n $(=2)$ and $l(=1)$. Note that the angular variable θ has disappeared from the final result! The probability density for the $2p$ state as a whole depends only on r and thus has the spherical symmetry that we expect it to have.

We now find the radial probability density for the $2p$ state as a whole, proceeding as we did for the ground state in Section 47-5; namely,

$$\begin{aligned} P_{2p}(r) &= [\psi_{2p}^2(r)][4\pi r^2] \\ &= \frac{1}{24a_0^3}\, r^4\, e^{-r/a_0}. \qquad (47\text{-}38) \end{aligned}$$

Figure 47-15*b* shows a plot of this radial probability density. Note from the plot that the maximum of the distribution occurs at $r = 4a_0$.

47-8 COUNTING THE STATES OF HYDROGEN

So far we have seen that there are really two "first excited states" of hydrogen, which we have labeled the $2s$ ($n = 2$, $l = 0$) and $2p$ ($n = 2$, $l = 1$). These states represent very different probability densities for the electron, but they have the same energy (E_2 in Eq. 47-21). Moreover, within the $2p$ state we have seen that there are three substates, corresponding to different values of the magnetic quantum number and thus again to different electron probability distributions. In this section we summarize our discussion of the hydrogen atom by arranging the excited states in an orderly sequence. However, we first must introduce one additional quantum number that is necessary to label the states.

Electron Spin

Most of the observed spectra of hydrogen and of the atoms beyond hydrogen in the periodic table could be well explained in terms of the principal quantum number, the orbital quantum number, and the magnetic quantum number. However, certain problems remained that could not be so explained. To mention just two of many: (1) The intense yellow light emitted by the sodium atom consists, not of a single wavelength, but of two closely spaced wavelengths. (2) Certain spectrum lines split into an array of closely spaced lines when the source is placed in a strong magnetic field. In 1924, Pauli pointed out the need for a fourth quantum number, associated intrinsically with the electron. A year later two Dutch graduate students, George Uhlenbeck (1900–1988) and Samuel Goudsmit (1902–1978), suggested that the electron, in addition to its orbital angular momentum in atoms, has an intrinsic angular momentum, called *spin angular momentum.*

The name "spin" conjures up an image of the electron as a tiny sphere spinning about an axis, much as the Earth spins about its axis of rotation. However, spin is a purely quantum phenomenon and this classical model is not valid. Spin is a property associated with all particles: electrons, protons, neutrons, photons, alpha particles, quarks—all have a spin angular momentum that describes their behavior.

The properties of the spin angular momentum in quantum mechanics are very similar to those of the orbital angular momentum. Each particle is assigned a spin quantum number s, usually called simply its spin. In contrast to the other quantum numbers we have discussed, the spin is an intrinsic property of the particle and does not depend on its state of motion. For example, the spin of the electron is 1/2. All electrons have $s = 1/2$, whether they are bound in atoms or moving as free particles. The spin is a fundamental property of the electron, like its mass or charge.

The intrinsic angular momentum is described by a vector $\vec{\mathbf{S}}$. In analogy with Eq. 47-28 for the orbital angular momentum, the magnitude or length of the spin angular momentum vector is

$$S = \sqrt{s(s+1)}\,\frac{h}{2\pi}. \tag{47-39}$$

Because $s = 1/2$ for all electrons, every electron is described by an angular momentum vector of length $\sqrt{3/4}\,(h/2\pi)$.

Again in analogy with the orbital angular momentum, we can specify the spatial orientation of the spin angular momentum by giving its z component. This component is given in terms of a *spin magnetic quantum number* m_s. Like the orbital magnetic quantum number m_l, the spin magnetic quantum number ranges from its minimum value $-s$ to its maximum value $+s$ in integer steps, so the only allowed values for m_s are

$$m_s = -1/2, +1/2. \tag{47-40}$$

The z component of the spin angular momentum is then

$$S_z = m_s\frac{h}{2\pi}. \tag{47-41}$$

The complete description of an electron in an atom then requires four quantum numbers: n, l, m_l, and m_s. The first three all arise in a natural way from solving Schrödinger's equation for the hydrogen atom. The spin quantum number is *not* predicted by Schrödinger's equation, but it does result from a more advanced calculation that combines quantum mechanics with relativity (Schrodinger's equation being a nonrelativistic equation).

In Chapter 48 we discuss the experimental evidence for electron spin, based on the properties of atoms in magnetic fields. In this section, we focus our attention on how our description of the states in hydrogen must be modified to take electron spin into account.

The States of Atomic Hydrogen

We have seen that the quantum states in which the hydrogen atom can exist are determined by four quantum numbers: n, l, m_l, and m_s. Let us see how, by using these numbers and the connections among them, we can arrange the hydrogen atom states in an orderly fashion.

The assembly of all hydrogen-atom states with the same principal quantum number n are said to form a *shell.* As Eq. 47-21 reveals, all such states have the same energy E_n. The collection of all states with the same value of the orbital angular momentum quantum number l is called a *subshell.* As Eq. 47-28 reveals, all such states have the same (magnitude of) angular momentum L.

The number of states within a given subshell is defined by the number of allowed values of the orbital magnetic quantum number m_l. For a given l, Eq. 47-29 tells us that there are $2l + 1$ such values. The effect of the spin magnetic quantum number m_s is simply to double this number, each state being associated with either $m_s = +1/2$ or $m_s = -1/2$. Thus, a subshell with $l = 2$ contains $2(2l + 1) = 2(2 \times 2 + 1) = 10$ states. Table 47-2 summarizes the classification of the hydrogen-atom states up to $n = 3$.

Do the numbers of states in the shells (that is, 2, 8, and 18) in the bottom row of Table 47-2 seem familiar, perhaps from your chemistry class? They are the numbers of elements in the horizontal rows (periods) of the periodic table of the elements! As Appendix E shows, period 1 contains 2 elements, periods 2 and 3 contain 8 elements each, and periods 4 and 5 contain 18 elements each. In Section 48-4 we shall see in detail just how the order of the elements in the periodic table arises from quantum-mechanical principles.

SAMPLE PROBLEM 47-10. A certain shell has principal quantum number $n = 4$. (*a*) How many subshells does this shell contain? (*b*) What is the number of states in each subshell? (*c*) What is the total number of states in the shell?

Solution (*a*) If $n = 4$, we know from Eq. 47-27 that the allowed values of l are 0, 1, 2, and 3. This is a total of four values. Can you show that the number of subshells in a given shell is always equal to n?

(*b*) The number of states in a subshell is given by $2(2l + 1)$, the factor 2 coming from the two allowed values of the spin quantum

TABLE 47-2 Quantum States of the Hydrogen Atom (up to $n = 3$ only)

n	1	2		3		
l	0	0	1	0	1	2
m_l	0	0	0, ±1	0	0, ±1	0, ±1, ±2
m_s	±1/2	±1/2	±1/2	±1/2	±1/2	±1/2
Number of states in the subshells	2	2	6	2	6	10
Number of states in the shells	2	8		18		

number m_s. For the numbers of states in the four subshells that make up the $n = 4$ shell, we then have

l	$2(2l+1)$
0	2
1	6
2	10
3	14

(*c*) The total number of states in the $n = 4$ shell is found by adding up the numbers of states in the subshells that the shell contains. From the above table we have $2 + 6 + 10 + 14 = 32$. Note that 32 is the number of elements in period 6 of the periodic table of the elements. See Appendix E.

Can you show that the number of states in a shell defined by principle quantum number n is always equal to $2n^2$? This works out in the present case because $2 \times 4^2 = 32$, as we found by explicit counting.

MULTIPLE CHOICE

47-1 Electrons, Free and Bound

47-2 An Electron Trapped in a Potential Well

1. An electron is trapped in an infinite one-dimensional potential well of width L.

(*a*) What is the minimum possible energy of the electron?
(A) 0
(B) Between 0 and $h^2/8mL^2$
(C) $\approx h^2/8mL^2$, but it is not possible to find the exact value because of the uncertainty principle
(D) Exactly $h^2/8mL^2$

(*b*) What is the minimum possible outcome of a measurement of the magnitude of the momentum of the electron?
(A) 0
(B) Between 0 and $h/2L$
(C) $\approx h/2L$, but it is not possible to find the exact value because of the uncertainty principle
(D) Exactly $h/2L$

2. An electron trapped in an infinite potential well has an energy of E. Which of the following statements about the width L of the well is most true?
(A) L must be greater than or equal to $(h^2/8mE)^{1/2}$.
(B) L must be less than or equal to $(h^2/8mE)^{1/2}$.
(C) L must be exactly equal to $(h^2/8mE)^{1/2}$.
(D) L can have any value.

3. An electron in an infinite potential well of width L_1 is in the ground state. An electron in a different infinite potential well of width L_2 is in the first excited state. The energy of the two electrons is the same. What is the ratio L_2/L_1?
(A) 4 (B) 3 (C) 2 (D) $\sqrt{2}$

47-3 An Electron Trapped in a Finite Well

4. Consider an electron trapped in a one-dimensional finite well of width L.

(*a*) What is the minimum possible kinetic energy of the electron?
(A) 0
(B) Between 0 and $h^2/8mL^2$
(C) $\approx h^2/8mL^2$, but it is not possible to find the exact value because of the uncertainty principle
(D) Exactly $h^2/8mL^2$

(*b*) What is the minimum possible outcome of a measurement of the magnitude of the momentum of the electron?
(A) 0
(B) Between 0 and $h/2L$
(C) $\approx h/2L$, but it is not possible to find the exact value because of the uncertainty principle
(D) Exactly $h/2L$

47-4 An Electron Trapped in an Atom

5. The presentation of the Bohr theory in the text was done under the assumption that an electron with a charge of magnitude $|e|$ "orbits" nuclear charge of magnitude $|e|$. By what factor would the energy levels change for an electron around a nuclear charge of magnitude $|Ze|$?
(A) Z (B) Z^2 (C) Z^4 (D) Z^{-2}

47-5 The Ground State of the Hydrogen Atom

47-6 Angular Momentum of Electrons in Atoms

6. An electron is in a spherically symmetric potential with $m_l = +2$. Which of the values for l are possible?
(A) l can have any integer value.
(B) l can have any integer value greater than 1.
(C) l can have any integer value less than 3.
(D) l must be equal to 2.

7. Consider the relationship between L_z and the magnitude of the angular momentum $L = |\vec{\mathbf{L}}|$.

(*a*) Can L_z be *exactly* equal to 0?
(A) Yes
(B) Yes, but only if $L = 0$
(C) No, because of the Heisenberg uncertainty principle
(D) No, because $|L_z|$ gets larger as L gets larger

(*b*) Can L_z be *exactly* equal to L?
(A) Yes
(B) No, and $|L_z|$ is always less than L
(C) No, and $|L_z|$ is always greater than L
(D) No, and $|L_z|$ could be larger or smaller than L

8. The magnitude of the orbital angular momentum of an electron is given by $L = \sqrt{5}\, h/\pi$. How many different values of L_z are possible?
(A) 4 (B) 5 (C) 8 (D) 9 (E) 11

9. How many values of m_l are possible for the ground state of a hydrogen atom?
(A) 0 (B) 1 (C) 2 (D) 3
(E) The answer cannot be determined without knowledge of l.

10. An experimenter attempts to measure several characteristics of an electron in the ground state of hydrogen. Which of the following quantities would have well-defined constant values?

(A) The total energy E
(B) The angular momentum $\vec{\mathbf{L}}$
(C) The z component of the angular momentum L_z
(D) The magnitude of the momentum p

47-7 An Excited State of the Hydrogen Atom

11. Which of the following sets of quantum numbers n, l, and m_l would *not* be possible for an excited state of hydrogen? (Choose all that are appropriate.)

(A) $n = 4, l = 2, m_l = -3$
(B) $n = 3, l = 2, m_l = -2$
(C) $n = 2, l = 2, m_l = +1$
(D) $n = 17, l = 2, m_l = -2$

12. Which of the following wavefunctions ψ_{nlm_l} for an excited state of hydrogen is spherically symmetric? (Choose all that are appropriate.)

(A) ψ_{310} (B) ψ_{31-1} (C) ψ_{400} (D) ψ_{32+1}

47-8 Counting the States of Hydrogen

13. Under which condition could the z component of the spin of an electron be zero?

(A) If S_x is $h/4\pi$
(B) If S_y is $h/4\pi$
(C) If both (A) and (B) are true
(D) If the electron is placed in a properly designed magnetic field
(E) S_z will never be zero for an electron.

14. There are subatomic relatives of the proton that have spin magnitudes of $S = |\vec{\mathbf{S}}| = \sqrt{15}\, h/4\pi$.

(*a*) What is the maximum possible value of S_z for these particles?

(A) $h/4\pi$ (B) $h/2\pi$ (C) $3h/4\pi$ (D) $5h/4\pi$

(*b*) How many different values of S_z are possible for these particles?

(A) 1 (B) 2 (C) 3 (D) 4 (E) 5

15. An electron is in a state in a hydrogen atom with quantum number $m_l = 2$.

(*a*) What can be said about the principal quantum number n for this electron?

(A) n is greater than 2.
(B) n is equal to 3.
(C) n is less than 3.
(D) Nothing can be said about n without more information.

(*b*) How many states exist with a lower energy level than the one occupied by this electron?

(A) Exactly 8 (B) Exactly 10
(C) At least 10 (D) At least 28

16. An electron is in a state in a hydrogen atom with quantum number $l = 3$.

(*a*) What can be said about the principal quantum number n for this electron?

(A) n is greater than 3.
(B) n is equal to 4.
(C) n is less than 4.
(D) Nothing can be said about n without more information.

(*b*) How many states exist with a lower energy level than the one occupied by this electron?

(A) Exactly 18 (B) Exactly 28
(C) At least 28 (D) At least 50

QUESTIONS

1. In Section 47-2 we solved the wave mechanical problem of a particle trapped in an infinitely deep well without ever using (or even writing down) Schrödinger's equation. How were we able to do that?

2. A standing wave can be viewed as the superposition of two traveling waves. Can you apply this view to the problem of a particle confined between rigid walls, giving an interpretation in terms of the motion of the particle?

3. The allowed energies for a particle confined between rigid walls are given by Eq. 47-3. First, convince yourself that, as n increases, the energy levels become farther apart. How can this possibly be? The correspondence principle would seem to require that they move closer together as n increases, approaching a continuum.

4. In the $n = 1$ state, for a particle confined between rigid walls, what is the probability that the particle will be found in a small length element at the surface of either wall?

5. What are the dimensions of $P_n(x)$ in Fig. 47-4? What is the value of the classically expected probability density? What value do the areas under the curves have? All these questions can be answered by inspection of the figure.

6. In Fig. 47-4 what do you imagine the curve for $P_n(x)$ for $n = 100$ looks like? Convince yourself that these curves approach classical expectations as $n \to \infty$.

7. Figure 47-4 shows that for $n = 3$ the probability function $P_n(x)$ for a particle confined between rigid walls is zero at two points between the walls. How can the particle ever move across these positions? (Hint: Consider the implications of the uncertainty principle.)

8. In Sample Problem 47-1, the electron's energy is determined exactly by the size of the box. How do you reconcile this with the fact that the uncertainty in the location of the electron cannot exceed 100 pm and, if the uncertainty principle is to be obeyed, the electron's momentum must be correspondingly uncertain?

9. Why was the Balmer series, rather than the Lyman or Paschen series, the first to be detected and analyzed in the hydrogen spectrum?

10. Any series of atomic hydrogen yet to be observed will probably be found in what region of the spectrum?

11. Can a hydrogen atom absorb a photon whose energy exceeds its binding energy (13.6 eV)?

12. On emitting a photon, an isolated hydrogen atom recoils to conserve momentum. Explain the fact that the energy of the emitted photon is slightly less than the energy difference between the energy levels involved in the emission process.

13. Radio astronomers observe lines in the hydrogen spectrum that originate in hydrogen atoms that are in states with $n = 350$ or so. Why can hydrogen atoms in states with such high quantum numbers not be produced and studied in the laboratory?

14. Only a relatively small number of Balmer lines can be observed from laboratory discharge tubes, whereas a large number are observed in stellar spectra. Explain this in terms of the small density, high temperature, and large volume of gases in stellar atmospheres.

15. Why does the concept of Bohr orbits violate the uncertainty principle?

16. Consider a hydrogen-like atom in which a positron (a positively charged electron) circulates about a (negatively charged) antiproton. In what way, if any, would the emission spectrum of this "antimatter atom" differ from the spectrum of a normal hydrogen atom?

17. If Bohr's theory and Schrödinger's wave mechanics predict the same result for the energies of the hydrogen atom states, then why do we need wave mechanics, with its greater complexity?

18. Compare Bohr's theory and wave mechanics. In what respects do they agree? In what respects do they differ?

19. Why do we not observe space quantization for a spinning top?

20. The angular momentum of the electron in the hydrogen atom is quantized. Why is the linear momentum not also quantized? (Hint: Consider the implications of the uncertainty principle.)

21. Angular momentum is a vector and you might expect that it would take three quantum numbers to describe it, corresponding to the three space components of a vector. Instead, in an atom, only two quantum numbers characterize the angular momentum. Explain why.

22. Discuss how good an analogy the rotating Earth revolving about the Sun is to a spinning electron moving about a proton in the hydrogen atom.

23. An atom in a state with zero angular momentum has spherical symmetry as far as its interaction with other atoms is concerned. It is sometimes called a "billiard-ball atom." Explain.

24. Define and distinguish among the terms wave function, probability density, and radial probability density.

25. What are the dimensions and the SI units of a wave function, a probability density, and a radial probability density? Are the dimensions what you expect?

26. We assert that the number of quantum numbers needed for a complete description of the motion of the electron in the hydrogen atom is equal to the number of degrees of freedom that the electron possesses. What is this number? How can you justify it?

27. How can you account for the fact that in the state of the hydrogen atom with $n = 2$ and $l = 0$, the probability density is a maximum at $r = 0$ but the radial probability density is zero there? See Fig. 47-14.

28. Figure 47-15*a* shows the three probability densities for the hydrogen atom states with $n = 2$ and $l = 1$. What determines the direction in space that we choose for the z axis?

29. Use Heisenberg's uncertainty principle to show that the probability densities in an $l = 2$ state have cylindrical symmetry about the z axis.

EXERCISES

47-1 Electrons, Free and Bound

47-2 An Electron Trapped in a Potential Well

1. (*a*) Calculate the smallest allowed energy of an electron confined to an infinitely deep well with a width equal to the diameter of an atomic nucleus (about 1.4×10^{-14} m). (*b*) Repeat for a neutron. (*c*) Compare these results with the binding energy (several MeV) of protons and neutrons inside the nucleus. On this basis, should we expect to find electrons inside nuclei?

2. What must be the width of an infinite well such that a trapped electron in the $n = 3$ state has an energy of 4.70 eV?

3. An electron, trapped in an infinite well of width 253 pm, is in the ground ($n = 1$) state. How much energy must it absorb to jump up to the third excited ($n = 4$) state?

4. The ground-state energy of an electron in an infinite well is 2.6 eV. What will the ground-state energy be if the width of the well is doubled?

5. (*a*) Calculate the separation in energy between the lowest two energy levels for a container 20 cm on a side containing argon atoms. (*b*) Find the ratio with the thermal energy of the argon atoms at 300 K. (*c*) At what temperature does the thermal energy equal the spacing between these two energy levels? Assume, for simplicity, that the argon atoms are trapped in a one-dimensional well 20 cm wide. The molar mass of argon is 39.9 g/mol.

6. (*a*) Calculate the fractional difference between two adjacent energy levels of a particle confined in a one-dimensional well of infinite depth. (*b*) Discuss the result in terms of the correspondence principle.

7. Consider an electron trapped in an infinite well whose width is 98.5 pm. If it is in a state with $n = 15$, what are (*a*) its energy? (*b*) The uncertainty in its momentum? (*c*) The uncertainty in its position?

8. Repeat Sample Problem 47-4, but assume now that the electron is in the $n = 2$ state.

9. Where are the points of (*a*) maximum and (*b*) minimum probability for a particle trapped in an infinitely deep well of length L if the particle is in the state n?

47-3 An Electron Trapped in a Finite Well

47-4 An Electron Trapped in an Atom

10. (*a*) Find the wavelength of the least energetic photon in the Lyman series. (*b*) Find the wavelength of the series limit of the Lyman series.

11. By direct substitution of numerical values of the fundamental constants, verify that the energy of the ground state of the hydrogen atom is -13.6 eV; see Eq. 47-21.

12. What are the (*a*) wavelength, (*b*) momentum, and (*c*) energy of the photon that is emitted when a hydrogen atom undergoes a transition from the state $n = 3$ to $n = 1$?

13. Calculate the five longest wavelengths of the Balmer series.

14. Show, on an energy-level diagram for hydrogen, the quantum numbers corresponding to a transition in which the wavelength of the emitted photon is 121.6 nm.

15. Calculate the binding energy (the energy necessary to remove the electron) of the hydrogen atom in the first excited state.

16. Find the value of the quantum number for a hydrogen atom that has an orbital radius of 847 pm.

17. Light with a wavelength of 1281.8 nm is emitted by a hydrogen atom. (*a*) What transition of the hydrogen atom is responsible for this radiation? (*b*) To what series does this radiation belong? (Hint: See Fig. 47-10.)

18. A hydrogen atom is excited from a state with $n = 1$ to one with $n = 4$. (*a*) Calculate the energy that must be absorbed by the atom. (*b*) Calculate and display on an energy-level diagram the different photon energies that may be emitted if the atom returns to the $n = 1$ state.

19. The lifetime of an electron in the state $n = 2$ in hydrogen is about 10 ns. What is the uncertainty in the energy of the $n = 2$ state? Compare this with the energy of this state.

20. (*a*) Calculate, according to the Bohr model, the speed of the electron in the ground state of the hydrogen atom. (*b*) Calculate the corresponding de Broglie wavelength. (*c*) Comparing the answers to (*a*) and (*b*), find a relation between the de Broglie wavelength λ and the radius a_0 of the ground-state Bohr orbit.

21. A neutron, with kinetic energy of 6.0 eV, collides with a resting hydrogen atom in its ground state. Show that this collision must be elastic (that is, energy must be conserved). (Hint: Show that the atom cannot be raised to a higher excitation state as a result of the collision.)

22. A hydrogen atom in a state having a binding energy (the energy required to remove an electron) of 0.85 eV makes a transition to a state with an excitation energy (the difference in energy between the state and the ground state) of 10.2 eV. (*a*) Find the energy of the emitted photon. (*b*) Show this transition on an energy-level diagram for hydrogen, labeling it with the appropriate quantum numbers.

23. According to the correspondence principle, as $n \rightarrow \infty$ we expect classical results in the Bohr atom. Hence the de Broglie wavelength associated with the electron (a quantum result) should get smaller compared with the radius of the Bohr orbit as n increases. Indeed, we expect that $\lambda/r \rightarrow 0$ as $n \rightarrow \infty$. Show that this is the case.

24. Calculate the recoil speed of a hydrogen atom, assumed to be initially at rest, if the electron makes a transition from the $n = 4$ state directly to the ground state. (Hint: Apply conservation of linear momentum.)

25. From the energy-level diagram for hydrogen, explain the observation that the frequency of the second Lyman-series line is the sum of the frequencies of the first Lyman-series line and the first Balmer-series line. This is an example of the empirically discovered *Ritz combination principle.* Use the diagram to find some other valid combinations.

26. (*a*) How much energy is required to remove the electron from a He^+ ion in its ground state? (*b*) From a Li^{2+} ion in a state with $n = 3$?

27. In stars the Pickering series is found in the He^+ spectrum. It is emitted when the electron in He^+ jumps from higher levels to the level with $n = 4$. (*a*) Show that the wavelengths of the lines in this series are given by

$$\lambda = \frac{Cn^2}{n^2 - 16},$$

in which $n = 5, 6, 7, \ldots$, and find the value of the constant C. (*b*) Calculate the wavelength of the first line in this series and of the series limit. (*c*) In what region(s) of the spectrum does this series occur?

28. In the ground state of the hydrogen atom, according to Bohr's theory, what are (*a*) the quantum number, (*b*) the orbit radius, (*c*) the angular momentum, (*d*) the linear momentum, (*e*) the angular velocity, (*f*) the linear speed, (*g*) the force on the electron, (*h*) the acceleration of the electron, (*i*) the kinetic energy, (*j*) the potential energy, and (*k*) the total energy?

29. How do the quantities (*b*) to (*k*) in Problem 28 vary with the quantum number n?

30. Suppose that we wish to test the possibility that electrons in atoms move in orbits by "viewing" them with photons with sufficiently short wavelength—say, 10.0 pm. (*a*) What would be the energy of such photons? (*b*) How much energy would such a photon transfer to a free electron in a head-on Compton collision? (*c*) What does this tell you about the possibility of confirming orbital motion by "viewing" an atomic electron at two or more points along its path? Assume that the speed of the electron is $0.10c$.

31. (*a*) Calculate the wavelength intervals over which the Lyman, the Balmer, and the Paschen series extend. (The interval extends from the longest wavelength to the series limit.) (*b*) Find the corresponding frequency intervals.

47-5 The Ground State of the Hydrogen Atom

32. In Fig. 47-12, verify the plotted values of $P(r)$ at (*a*) $r = 0$, (*b*) $r = a_0$, and (*c*) $r = 2a_0$.

33. In the ground state of the hydrogen atom, evaluate the square of the wave function $\psi^2(r)$ and the radial probability density $P(r)$ for the positions (*a*) $r = 0$ and (*b*) $r = a_0$. Explain what these quantities mean.

34. A spherical region of radius $0.05a_0$ is located a distance a_0 from the nucleus of a hydrogen atom in its ground state. Calculate the probability that the electron will be found inside this sphere. (Assume that ψ is constant inside the sphere.)

35. Find the ratio of the probabilities of finding the electron in the hydrogen atom in a thin shell at the Bohr radius to that of finding it in a shell of the same thickness at twice that distance.

36. For a hydrogen atom in its ground state, calculate the probability of finding the electron between two spheres of radii $r = 1.000a_0$ and $r = 1.016a_0$.

47-6 Angular Momentum of Electrons in Atoms

37. For a hydrogen atom in a state with $l = 3$, calculate the allowed values of (*a*) L_z and (*b*) θ. (*c*) Find also the magnitude of $\vec{\mathbf{L}}$. Where appropriate, express answers in units of h.

38. If an electron in a hydrogen atom is in a state with $l = 5$, what is the smallest possible angle between $\vec{\mathbf{L}}$ and L_z?

39. Show that Eq. 47-33 is a plausible version of the uncertainty principle $\Delta p \cdot \Delta x = h/2\pi$. (Hint: Multiply by r/r; associate p with mv, and L with mvr.)

47-7 An Excited State of the Hydrogen Atom

40. Using Eq. 47-38, show that, for the hydrogen atom state with $n = 2$ and $l = 1$,

$$\int_0^\infty P(r)\,dr = 1.$$

What is the physical interpretation of this result?

41. For the state $n = 2$, $l = 0$, (*a*) locate the two maxima for the radial probability density curve of Fig. 47-14*b*, and (*b*) calculate the values of the radial probability density at the two maxima; compare with Fig. 47-14*b*.

42. For a hydrogen atom in a state with $n = 2$ and $l = 0$, calculate the probability of finding the electron between two spheres of radii $r = 5.00a_0$ and $r = 5.01a_0$.

47-8 Counting the States of Hydrogen

43. Write down the quantum numbers for all the hydrogen atom states belonging to the subshell for which $n = 4$ and $l = 3$.

44. A hydrogen atom state is known to have the quantum number $l = 3$. What are the possible n, m_l, and m_s quantum numbers?

45. A hydrogen atom state has a maximum m_l value of $+4$. What can you say about the rest of its quantum numbers?

46. How many hydrogen atom states are there with $n = 5$? How are they distributed among the subshells?

47. What are the quantum numbers n, l, m_l, m_s for the two electrons of the helium atom in its ground state?

48. Calculate the two possible angles between the electron spin angular momentum vector and the z axis.

49. Label as true or false these statements involving the quantum numbers n, l, m_l. (*a*) One of these subshells cannot exist: $n = 2$, $l = 1$; $n = 4$, $l = 3$; $n = 3$, $l = 2$; $n = 1$, $l = 1$. (*b*) The number of values of m_l that are allowed depends only on l and not on n. (*c*) The $n = 4$ shell contains four subshells. (*d*) The smallest value of n that can go with a given l is $l + 1$. (*e*) All states with $l = 0$ also have $m_l = 0$, regardless of the value of n. (*f*) Every shell contains n subshells.

50. Show that the number of states in any shell is given by $2n^2$.

PROBLEMS

1. Consider a conduction electron in a cubical crystal of a conducting material. Such an electron is free to move throughout the volume of the crystal but cannot escape to the outside. It is trapped in a three-dimensional infinite well. The electron can move in three dimensions, so that its total energy is given by (compare with Eq. 47-3),

$$E = \frac{h^2}{8mL^2}(n_1^2 + n_2^2 + n_3^2),$$

in which n_1, n_2, n_3 each take on the values 1, 2,. . . . Calculate the energies of the lowest five distinct states for a conduction electron moving in a cubical crystal of edge length $L = 250$ nm.

2. A particle is confined between rigid walls separated by a distance L. (*a*) Show that the probability P that it will be found within a distance $L/3$ from one wall is given by

$$P = \frac{1}{3}\left(1 - \frac{\sin(2\pi n/3)}{2\pi n/3}\right).$$

Evaluate the probability for (*b*) $n = 1$, (*c*) $n = 2$, (*d*) $n = 3$, and (*e*) under the assumption of classical physics.

3. A particle is confined between rigid walls located at $x = 0$ and $x = L$. For the $n = 4$ energy state, (*a*) sketch the probability density curve for the particle's location. Calculate the approximate probabilities of finding the particle within a region $\Delta x = 0.0003L$ when (*b*) Δx is located at $x = L/8$ and (*c*) at $x = 3L/16$. Refer to your figure to see whether or not your results seem reasonable. (Hint: No integration is necessary.)

4. The wave function for an electron in a certain one-dimensional potential is $\Psi_0(x,t) = A_0 e^{-\pi m\omega x^2/h - i\omega t/2}$ where A_0 is a real number. (*a*) Show that the probability density is real. (*b*) Find A_0 assuming that the total probability of finding the electron anywhere is 1. (Hint: Use integral 17 in Appendix I.) (*c*) Where on the x axis is the electron most likely to be found?

5. The time-independent portion of a wave function for an electron of energy $E = h\omega/4\pi$ is $\psi_0(x) = A_0 e^{-\pi m\omega x^2/h}$, where A_0 is a constant. Substitute this into Schrödinger's equation (Eq. 46-17) and find the equation for the potential energy $U(x)$. What does this potential energy represent?

6. If an electron is revolving in an orbit at frequency f_0, classical electromagnetism predicts that it will radiate energy not only at this frequency but also at $2f_0$, $3f_0$, $4f_0$, and so on. Show that this is also predicted by Bohr's theory of the hydrogen atom in the limiting case of large quantum numbers.

7. Apply Bohr's model to a hydrogen-muon atom, which consists of a nucleus of charge e with a negative muon (an elementary particle with a charge $q = -e$ and a mass $m = 207m_e$, where m_e is the electron mass) circulating about it. Calculate (*a*) the muon-nucleus separation in the first Bohr orbit, (*b*) the ionization energy, and (*c*) the wavelength of the most energetic photon that can be emitted. See "The Muonium Atom," by Vernon W. Hughes, *Scientific American,* April 1966, p. 93.

8. Apply Bohr's model to the positronium atom. This consists of a positive and a negative electron revolving around their center of mass, which lies halfway between them. (*a*) What relationship exists between this spectrum and the hydrogen spectrum? (*b*) What is the radius of the ground-state orbit? (Hint: Calculate the reduced mass of the atom.) See "Exotic Atoms," by E. H. S. Burhop, *Contemporary Physics,* July 1970, p. 335.

9. Of the three components of $\vec{\mathbf{L}}$, one, L_z, is quantized, according to Eq. 47-30. In view of the restrictions imposed by Eqs. 47-27 and 47-28, taken together, show that the most that can be said about the other two components of $\vec{\mathbf{L}}$ is

$$\sqrt{L_x^2 + L_y^2} = \sqrt{l(l+1) - m_l^2}\, h/2\pi.$$

Note that these two components are not separately quantized. Show also that

$$\sqrt{l}\, h/2\pi \leq \sqrt{L_x^2 + L_y^2} \leq \sqrt{l(l+1)}\, h/2\pi.$$

10. In atoms there is a finite, though very small, probability that, at some instant, an orbital electron will actually be found inside the nucleus. In fact, some unstable nuclei use this occasional appearance of the electron to decay by *electron capture.* Assuming that the proton itself is a sphere of radius 1.1×10^{-15} m and that the hydrogen atom electron wave function holds all the way to the proton's center, use the ground-state wave function to calculate the probability that the hydrogen atom electron is inside its nucleus. (Hint: When $x \ll 1$, $e^{-x} \approx 1$.)

11. Repeat Problem 10 for an electron in the state $n = 2$, $l = 0$; that is, calculate the probability that the electron will be found inside the proton, radius = 1.1 fm, that constitutes the nucleus of the hydrogen atom.

12. (*a*) In the ground state of the hydrogen atom, show that the probability P that the electron lies within a sphere of radius r is given by

$$P = 1 - e^{-2x}(1 + 2x + 2x^2),$$

in which $x = r/a_0$. (*b*) Evaluate the probability that, in the ground state, the electron lies within a sphere of radius a_0.

13. Use the result of Problem 12 to calculate the probability that the electron in a hydrogen atom, in the ground state, will be found between the spheres $r = a_0$ and $r = 2a_0$.

14. For an electron in the ground state of the hydrogen atom, calculate the radius of a sphere for which the probability that the electron will be found inside the sphere equals the probability that the electron will be found outside the sphere. (Hint: See Problem 12.)

15. For a hydrogen atom in a state with $n = 2$ and $l = 0$, what is the probability of finding the electron somewhere within the smaller of the two maxima of its radial probability density function? See Fig. 47-14*b*.

COMPUTER PROBLEM

1. The finite square well problem can be solved numerically by setting up a spreadsheet similar to

x (pm)	ψ	$d\psi/dx$	$d^2\psi/dx^2$
0	A	B	
1			
2			
etc.			

where for each new row ψ is given by the previous row's $\psi + \Delta x(d\psi/dx)$, similarly $d\psi/dx$ is given by the previous row's $d\psi/dx + \Delta x(d^2\psi/dx^2)$. However, $d^2\psi/dx^2$ is given by Schrödinger's equation in terms of the energy E, position x, and wave function ψ. L is the width of the well; the iterations are started from the center of the well because then either A or B, the initial values of ψ or $d\psi/dx$, is zero depending on whether the state is even or odd. (*a*) Consider an electron bound in the well of width 100 pm and depth 250 eV. Find the ground state by choosing the minimum value for the energy E that prevents ψ from "blowing up" before reaching $x = 100$ pm on the spreadsheet. Remember that x is measured from the center of the well. The accuracy of your estimate for E can be improved by requiring that ψ remain finite for larger values of x. For the ground state, set $A = 1$ and $B = 0$. (*b*) Find the energy of the first excited state. In this case, set $A = 0$ and $B = 1$; find the lowest energy E that prevents ψ from "blowing up" inside some specified distance.

CHAPTER 48

ATOMIC STRUCTURE

In this chapter we turn to the study of multielectron atoms and discuss the motions of electrons that lie deep within the atom and are responsible for the production of x rays. X-ray studies then lead to the development of the concept of atomic number.

Next we consider the quantum rules for constructing multielectron atoms and we see how those rules can be used to determine the arrangement of the elements in the periodic table. We then describe a number of experiments involving the properties of atoms and we show that these results can be understood in terms of the quantum theory of the atom. Finally, we use information from atomic structure to analyze the operation of a helium–neon laser.

48-1 THE X-RAY SPECTRUM OF ATOMS

So far we have focused our attention on the motions of the single electron in the hydrogen atom. We now consider atoms with more than one electron and we examine the motions of electrons that lie deep within such atoms. We move from a region of relatively low binding energy (5 eV for the work required to remove the valence electron from sodium, for example) to a region of higher binding energy (70 keV to remove an innermost electron from tungsten, for example). Although still part of the electromagnetic spectrum, the radiations emitted by such deep-seated electrons differ drastically in wavelength—for example, from $\sim 6 \times 10^{-7}$ m for the sodium doublet lines to $\sim 2 \times 10^{-11}$ m for one of the tungsten characteristic radiations, a ratio of about 3×10^4. Radiation of such short wavelength falls in the x-ray region of the spectrum. Figure 48-1 shows the wavelength spectrum of the x rays produced when 35-keV electrons are allowed to strike a molybdenum target.

The Continuous X-ray Spectrum

We first examine the continuous spectrum of Fig. 48-1, ignoring—for the time being—the two prominent peaks that rise above it. Consider an electron of kinetic energy K that passes close to the nucleus of one of the molybdenum atoms in the target, as in Fig. 48-2. In such a collision, the electron loses kinetic energy, which appears as the energy hf of an

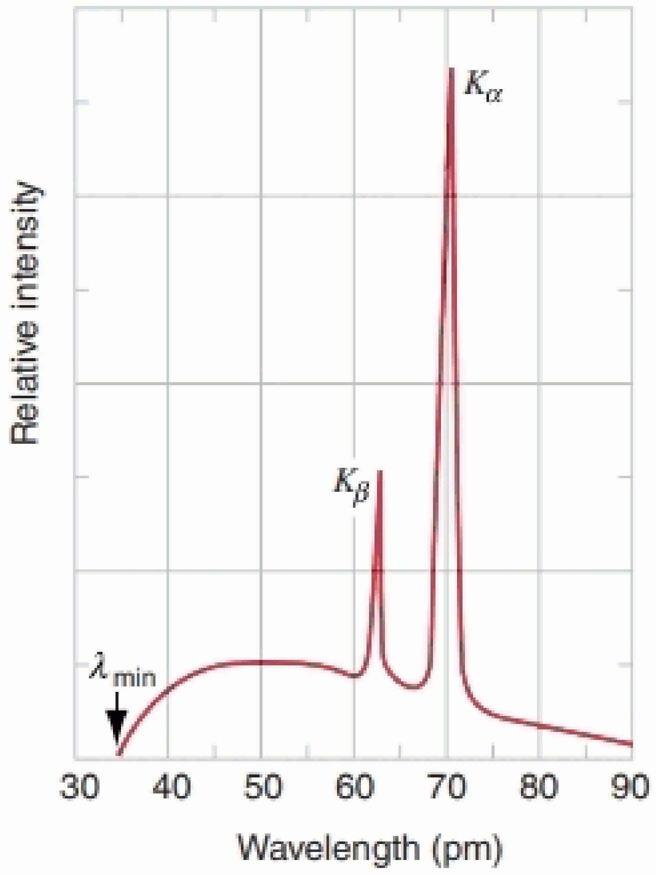


FIGURE 48-1. The distribution of wavelengths that results when a molybdenum target is bombarded with 35-keV electrons. Note the sharply defined minimum wavelength.

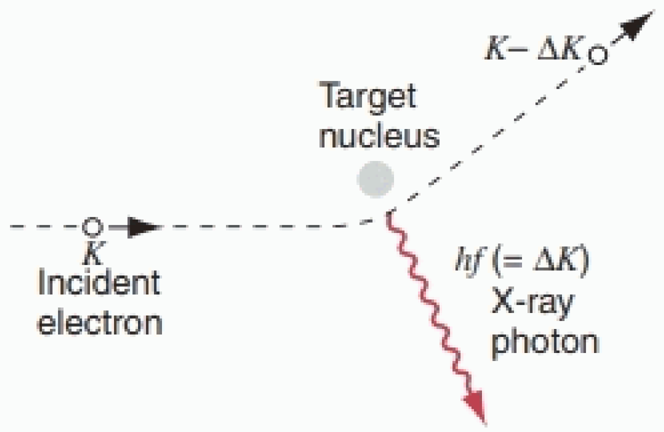


FIGURE 48-2. The continuous x-ray spectrum consists of photons generated when incident electrons pass near individual target nuclei, losing energy in the process.

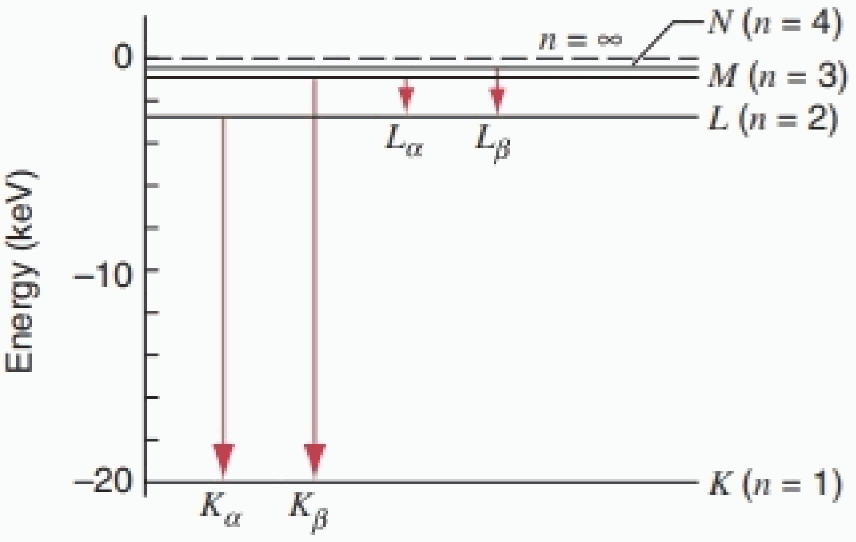


FIGURE 48-3. An atomic energy level diagram for molybdenum, showing the transitions that represent the characteristic x-ray spectrum of that element. All levels except the K level consist of a number of close-lying components, which are not shown in the figure.

x-ray photon that radiates away from the site of the encounter. (The energy acquired by the relatively massive molybdenum nucleus is so small that it may be ignored.) This radiation is called *bremsstrahlung* (German for "braking radiation") and it accounts for the continuous x-ray spectrum.

A prominent feature of the continuous spectrum of Fig. 48-1 is the sharply defined cutoff wavelength $\lambda_{\min}$, below which the continuous spectrum does not exist. This minimum wavelength corresponds to an event in which one of the incident electrons (with kinetic energy $e\,\Delta V$, acquired as a result of acceleration through a potential difference ΔV) loses *all* of its energy in a single encounter, radiating it away as a single photon. Thus

$$eV = hf_{\max} = \frac{hc}{\lambda_{\min}}$$

or

$$\lambda_{\min} = \frac{hc}{e\,\Delta V}. \tag{48-1}$$

Because the Planck constant h appears in Eq. 48-1, the very existence of a minimum wavelength is a quantum phenomenon.

If you change the target material, perhaps from molybdenum to copper, the general shape and intensity of the continuous spectrum will change but the cutoff wavelength will *not* change. It depends only on the kinetic energy of the electrons that bombard the target and not at all on the target material.

The Characteristic X-ray Spectrum

We now turn our attention to the two peaks of Fig. 48-1, labeled K_α and K_β. These peaks are characteristic of the target material and, together with other peaks that appear at longer wavelengths and are not shown in Fig. 48-1, form the *characteristic x-ray spectrum* of the target element.

Here is how these x-ray photons arise: (1) An energetic electron strikes a target atom and knocks out one of its deep-seated electrons. If the electron is in a shell with $n = 1$ (called, for historical reasons, the K shell) there remains a vacancy or a "hole," as we shall call it, in this shell. (2) One of the outer electrons moves in to fill this hole and, in the process, the atom emits a characteristic x-ray photon. If the electron falls from the shell with $n = 2$ (called the L shell) we have the K_α line of Fig. 48-1; if it falls from the next outermost shell (called the M shell) we have the K_β line, and so on. Of course, such a transition will leave a hole in the L shell or the M shell, but this will be filled by an electron from further out in the atom.

Figure 48-3 shows an energy level diagram for molybdenum, the element used as the target to obtain the spectrum shown in Fig. 48-1. The arrow marked K_α shows the result of an electron from the L ($n = 2$) shell dropping down to fill a vacancy in the K shell. The energies of these levels are $E_2 = -2.6$ keV and $E_1 = -20.0$ keV. The energy of the K_α photon emitted when the electron makes this jump is then $E = E_2 - E_1 = 17.4$ keV, typical for photons in the x-ray region of the electromagnetic spectrum. The wavelength corresponding to this photon energy is 71 pm, consistent with where the peak marked K_α appears in Fig. 48-1.

Instead of an electron from the L shell, one from the M shell can drop down to fill the K-shell vacancy. This jump is indicated by the arrow marked K_β in Fig. 48-3. Given the energy of the M ($n = 3$) level as -0.4 keV, you should be able to show that the wavelength of the K_β x-ray photon agrees with the corresponding peak in Fig. 48-1.

In the case of the K_α x ray, a hole will be left in the L shell when an electron jumps down to fill the K-shell vacancy. This hole can be filled by electrons dropping down from higher levels. Two of these jumps (from the M and N shells) are indicated in Fig. 48-3 as the L_α and L_β x rays. Can you explain why the L_α and L_β peaks do not appear in Fig. 48-1?

SAMPLE PROBLEM 48-1. Calculate the wavelength $\lambda_{\min}$ for the continuous spectrum of x rays emitted when 35-keV electrons fall on a molybdenum target, as in Fig. 48-1.

Solution From Eq. 48-1 we have

$$\lambda_{\min} = \frac{hc}{e\,\Delta V} = \frac{(6.63 \times 10^{-34}\ \text{J}\cdot\text{s})(3.00 \times 10^{8}\ \text{m/s})}{(35.0 \times 10^{3}\ \text{eV})(1.60 \times 10^{-19}\ \text{J/eV})}$$
$$= 3.55 \times 10^{-11}\ \text{m} = 35.5\ \text{pm}.$$

This agrees with the experimental result labeled $\lambda_{\min}$ in Fig. 48-1. (Note that the quantity in the second set of parentheses in the denominator is a conversion factor.)

48-2 X RAYS AND THE NUMBERING OF THE ELEMENTS

From the time of Mendeleev (1834–1907) the elements were listed in the periodic table in the order of increasing *atomic weight.* One difficulty was that, in several cases, the strict order of increasing atomic weights had to be reversed to maintain the similarity of chemical properties in the vertical columns of the table. The young British physicist H. G. J. Moseley, whose promising research career was cut short at age 27 by a sniper's bullet in World War I, first introduced the concept of *atomic number* (symbol Z) and showed that these difficulties were removed if the periodic table was organized in order of increasing Z. In effect, Moseley showed how to string the elements in a line and to assign place numbers to them. His method was based on the measurement and analysis of the characteristic x rays of the elements.

In his investigations, Moseley measured the wavelengths of the characteristic x rays of as many elements as he could find (he found 38), using them as targets for electron bombardment in an x-ray tube of his own design. He then sought, and readily found, regularities in the spectra as he moved from element to element in the periodic table. In particular he noted that if, for a given x-ray spectrum line such as K_α, he plotted the square root of its frequency ($= \sqrt{f} = \sqrt{c/\lambda}$) against the position of the associated element in the periodic table, a straight line resulted. Figure 48-4 shows a portion of his data. Mosely's conclusion from the full body of his data was:

> *We have here a proof that there is in the atom a fundamental quantity, which increases by regular steps as we pass from one element to the next. This quantity can only be the charge on the central atomic nucleus.*

Thus did Moseley discover the concept, and the physical significance, of atomic number.

Moseley's data, shown in Fig. 48-4, can be represented by the linear relation

$$\sqrt{f} = C(Z - 1), \tag{48-2}$$

in which C is a constant and Z is the numerical position of the element in the periodic table. This equation, which can be derived from Bohr's hybrid quantum theory, fits the data very well.

Moseley's achievement can be appreciated all the more when we realize the state of understanding of atomic structure at that time (1913). The nuclear model of the atom had been proposed by Rutherford only two years earlier. Little was known about the magnitude of the nuclear charge or about the arrangement of the atomic electrons. Quantum mechanics had yet to be discovered. The periodic table had several empty squares, and a surprisingly large number of claims for new elements had been advanced. The rare earth elements, because of the problems caused by their similar chemical properties, had not yet been properly sorted out. Due to Moseley's work, the characteristic x-ray spectrum became the universally accepted signature of an element.

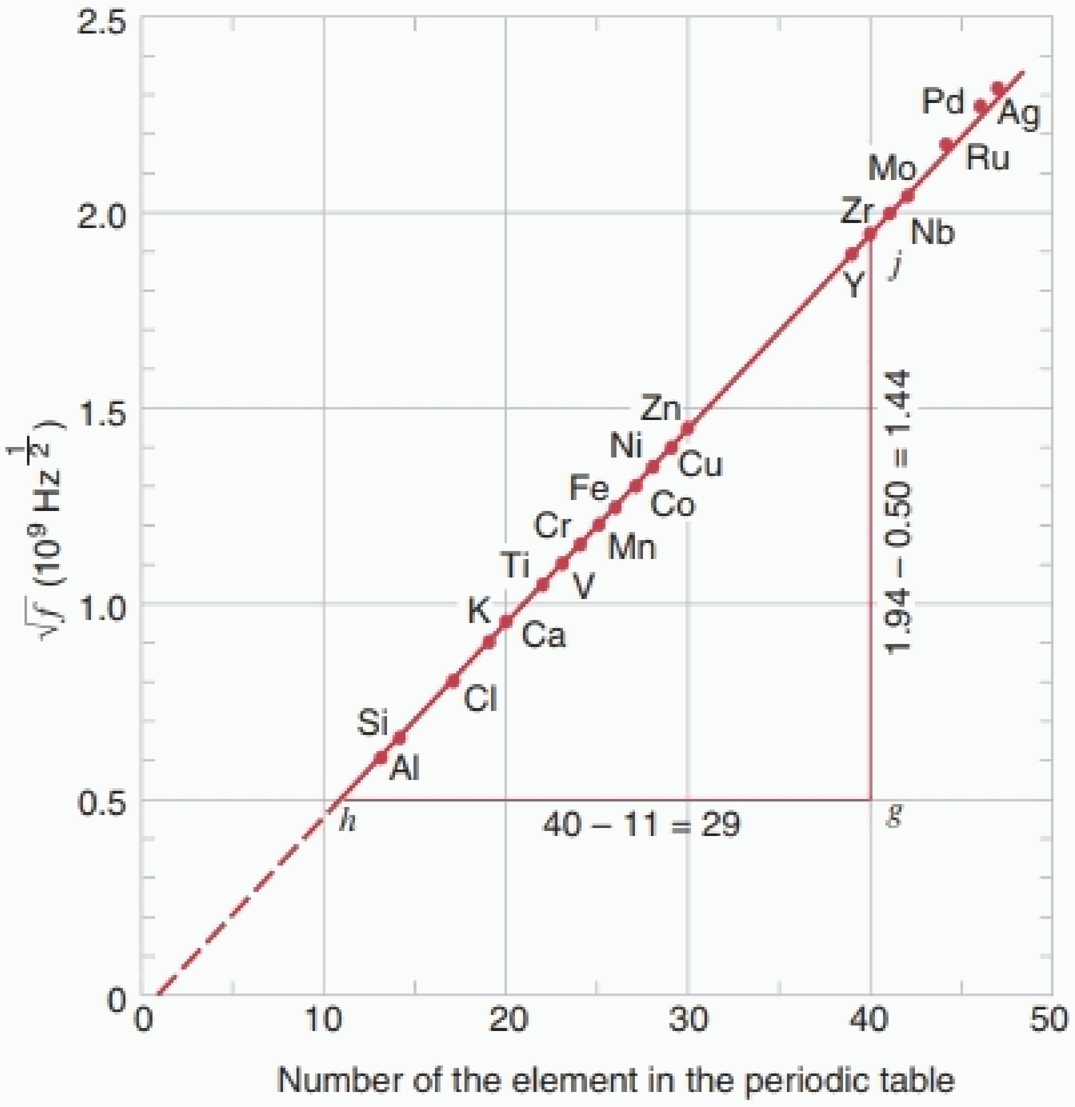


FIGURE 48-4. A Moseley plot of the K_α line of the x-ray spectrum of 21 elements. The frequency is determined from the measured wavelength.

It is not hard to see why the characteristic x-ray spectrum shows such remarkable regularities from element to element whereas the visible spectrum does not. The key to the identity of an element is the charge on its nucleus. This determines the number of its atomic electrons and thus its chemical properties. Gold, for example, is what it is because its atoms have a nuclear charge of $+79e$. If it had one more unit of charge, it would not be gold but mercury; if it had one less, it would be platinum. The K-shell electrons, which play such a large role in the generation of the characteristic x-ray spectra, lie very close to the nucleus and are sensitive probes of its charge. The visible spectrum, on the other hand, is associated with transitions of the outer electrons, which are heavily shielded or "screened" from the nucleus by the remaining atomic electrons; they are not sensitive probes of nuclear charge.

Bohr Theory and the Moseley Plot

Bohr's theory works well for hydrogen but fails for atoms with more than one electron (in part because it does not include the repulsive interaction between the electrons).

Nevertheless, it provides an excellent first approximation in accounting for the Moseley plot of Fig. 48-4.

Consider an electron in the L shell of an atom that is about to move into a hole in the K shell, emitting a K_α x-ray photon in the process. Using Gauss' law (see Eq. 27-16), we find that the electric field at the location of the L electron is determined by the charge enclosed in an imaginary sphere of radius equal to the radial coordinate of the L electron. This sphere encloses a charge $+Ze$ from the nucleus and a charge $-e$ from the single remaining K electron. We say that the K electron "screens" the charge of the nucleus. In part because of this screening and in part because of readjustments that take place in the electron cloud as a whole, the effective atomic number for the transition turns out to be $Z - b$, where $b \approx 1$.

Bohr's formula for the frequency of the radiation corresponding to a transition in a hydrogen-like atom between any two atomic levels differing in energy by ΔE is

$$f = \frac{\Delta E}{h} = \frac{mZ^2e^4}{8\epsilon_0^2h^3}\left(\frac{1}{n_\mathrm{f}^2} - \frac{1}{n_\mathrm{i}^2}\right), \tag{48-3}$$

using Eq. 47-23 for the energy levels. For the K_α transition of Fig. 48-3, we can replace Z by $Z - b$ and substitute 1 for n_f and 2 for n_i. Doing so yields

$$f = \frac{m(Z-b)^2e^4}{8\epsilon_0^2h^3}\left(\frac{1}{1^2} - \frac{1}{2^2}\right)$$

Taking the square root of each side leads to

$$\sqrt{f} = \left(\frac{3me^4}{32\epsilon_0^2h^3}\right)^{1/2}(Z-b), \tag{48-4}$$

which we can write in the form

$$\sqrt{f} = C\,(Z-1), \tag{48-5}$$

where C is the indicated constant and $b \approx 1$.

Equation 48-5 represents a straight line, in full agreement with the experimental data of Fig. 48-4. If this plot is extended to higher atomic numbers, however, it turns out to be not quite straight but somewhat concave upward. Nevertheless, the quantitative agreement with Bohr theory is surprisingly good, as Sample Problem 48-2 shows.

SAMPLE PROBLEM 48-2. Calculate the value of the quantity C in Eq. 48-5 and compare it with the measured slope of the straight line in Fig. 48-4.

Solution Comparing Eqs. 48-4 and 48-5 allows us to write

$$\begin{aligned} C &= \left(\frac{3me^4}{32\epsilon_0^2h^3}\right)^{1/2} \\ &= \frac{\sqrt{3}\,(9.11\times10^{-31}\ \mathrm{kg})^{1/2}(1.60\times10^{-19}\ \mathrm{C})^2}{4\sqrt{2}\,(8.85\times10^{-12}\ \mathrm{F/m})(6.63\times10^{-34}\ \mathrm{J\cdot s})^{3/2}} \\ &= 4.95\times10^7\ \mathrm{Hz}^{1/2}. \end{aligned}$$

Careful measurement of Fig. 48-4, using the triangle hgj, yields a slope equal to

$$\frac{gj}{hg} = \frac{(1.94-0.50)\times10^9\ \mathrm{Hz}^{1/2}}{(40-11)} = 4.96\times10^7\ \mathrm{Hz}^{1/2},$$

which is in agreement with the value predicted by Bohr theory within the uncertainty of the graphical measurement. Note also that the intercept in Fig. 48-4 is in fact close to 1, as expected from our screening argument.

The agreement with Bohr theory is not nearly as good for other lines in the x-ray spectrum, corresponding to the transitions of electrons farther from the nucleus; here we must rely on calculations based on quantum mechanics.

SAMPLE PROBLEM 48-3. A cobalt target is bombarded with electrons, and the wavelengths of its characteristic x-ray spectrum are measured. A second, fainter characteristic spectrum is also found, due to an impurity in the target. The wavelengths of the K_α lines are 178.9 pm (cobalt) and 143.5 pm (impurity). What is the impurity?

Solution Let us apply Eq. 48-2 both to cobalt and to the impurity X. Putting $f = c/\lambda$, we have

$$\sqrt{\frac{c}{\lambda_\mathrm{Co}}} = C(Z_\mathrm{Co} - 1) \quad \text{and} \quad \sqrt{\frac{c}{\lambda_X}} = C(Z_X - 1).$$

Dividing yields

$$\sqrt{\frac{\lambda_\mathrm{Co}}{\lambda_X}} = \frac{Z_X - 1}{Z_\mathrm{Co} - 1}$$

and substituting the given values, we obtain

$$\sqrt{\frac{178.9\ \mathrm{pm}}{143.5\ \mathrm{pm}}} = \frac{Z_X - 1}{27 - 1}.$$

Solving for the unknown, we find $Z_X = 30.0$. A glance at the periodic table identifies the impurity as zinc.

48-3 BUILDING ATOMS

In the preceding section we have seen how to string the elements into a continuous line, assigning an atomic number Z to each. Here we use the principles of quantum mechanics to break up this line into segments that correspond to the horizontal periods of the periodic table (see Appendix E).

Our efforts will meet with total success. Every detail of the periodic table can be accounted for, including (1) the numbers of elements in the seven horizontal periods of the periodic table, (2) the similarity of the chemical properties in the various vertical columns of that table, and (3) the existence of the lanthanide and the actinide series of elements, each crammed into a single square of the table. In short, quantum mechanics, supplemented by certain guiding principles, accounts for every feature of this table and thus for all of chemistry.

Our plan is to start with a nucleus of charge $+Ze$ and, by adding Z electrons to it, to construct a neutral atom in its ground state. Success follows if we are guided by the following three rules of atom building:

1. *The quantum number principle:* We can use the four quantum numbers (n, l, m_l, m_s) that we have established for the hydrogen atom to identify the quantum states of individual electrons in multielectron atoms. Furthermore, the restrictions among these quantum numbers, implied in Table 47-2, still hold.

2. *The Pauli exclusion principle:* This powerful principle, advanced in 1925 by the Austrian physicist Wolfgang Pauli (1900–1958), asserts that, in a multielectron atom, no two electrons can have the same set of quantum numbers. If this principle did not hold, all the electrons in an atom would pile up in its K shell, and chemistry as we know it would not exist, because all atoms would have the same chemical properties. Biochemical processes, even those responsible for human existence, would not be possible.

3. *The minimum energy principle:* As we fill up the available quantum states with electrons, the question arises: In what order should we fill them? The answer is: In adding an electron, put it in whatever vacant quantum state leads to an atom that is lowest in energy. If you do otherwise, the atom will not be in its ground state. The lowest energy subshell can be identified with the help of the following rule: For a given quantum number n in a multielectron atom, the order of the increasing energy of the subshells is the order of increasing l. (However, as we discuss in the next section, the states don't always fill in order of increasing n.)

Consider, for example, the shell defined by $n = 4$ in a neutral lead atom. (The valence electrons for lead have $n = 6$, so the $n = 4$ electrons lie somewhere in the middle of the smeared-out cloud of electrons surrounding the lead nucleus.) A shell with $n = 4$ has four subshells, corresponding to $l = 0$, 1, 2, and 3, which we label as $4s$, $4p$, $4d$, and $4f$, respectively.

The binding energies of electrons in these subshells can be found experimentally by measuring the characteristic x-ray spectra of the lead atom. Table 48-1 shows the results. The $4s$ state has the lowest energy of the four. That is, you must add an energy of 890 eV to remove an electron in a $4s$ state from a lead atom. At the other limit, an energy of only 140 eV is needed to remove an electron from the $4f$ state.

TABLE 48-1 The Energies of Electrons with $n = 4$ in a Lead Atom ($Z = 82$)

Principal Quantum Number n	*Orbital Quantum Number l*	*Energy (eV)*
4	0 ($4s$)	−890
4	1 ($4p$)	−710
4	2 ($4d$)	−420
4	3 ($4f$)	−140

If we were to examine the probability density plots for the four states listed in Table 48-1, we would find that, the lower the value of l, the more likely are the electrons to be found near the lead nucleus. Some close-in electrons "see" a higher effective nuclear charge and more tightly bound; they lie lower in energy, just as the minimum energy principle and the example of Table 48-1 show.

48-4 THE PERIODIC TABLE

Figure 48-5 shows how the periodic table is put together, using the rules for atom building that we described in the preceding section. Energy increases upward in this figure. States with the same l value have been displaced to the left and grouped into columns. The shaded blocks represent the horizontal periods of the periodic table.

The dependence of energy on l is a dominant feature of Fig. 48-5. Look, for example, at the sequence $4s$, $4p$, $4d$, and $4f$. Although these states have the same principal quantum number, they do not have the same energy. Their energy increases with l and, in fact, the $4f$ states are so high that they are above the $5s$ and the $5p$ states, which have a higher value of n.

By starting with hydrogen ($Z = 1$) in Fig. 48-5 and following the curved line, we can see how the seven horizontal periods of the periodic table are built up, each starting with an alkali metal and ending with an inert gas. The long sixth period, for example, starts with the alkali metal cesium ($Z = 55$) and ends with the inert gas radon ($Z = 86$). The order in which the subshells are filled, as the curved line indicates, is $6s$, $4f$, $5d$, and $6p$.

The sixth period contains a run of 14 elements ($Z = 57$ to $Z = 70$), listed separately at the bottom of the periodic table in Appendix E. These elements are the *rare earths,* or the *lanthanides,* after the element lanthanum that begins the series. Their chemical properties are so similar that they are grouped into a single square of the periodic table. This similarity arises because, while the $4f$ states are being filled deep within the electron cloud, an outer screen of one or two $6s$ electrons remains in place. It is these outermost electrons that determine the chemical properties of the atom. A similar series (the *actinides*) occurs in the seventh period, the outermost screening electrons in this case being in a $7s$ state.

The maximum number of electrons in a given subshell is $2(2l + 1)$. If you count the number of elements in each of the subshells of Fig. 48-5, you will find them consistent with this rule. That is, 2 for s states ($l = 0$), 6 for p states ($l = 1$), 10 for d states ($l = 2$), and 14 for f states ($l = 3$).

Electron Configurations of the Elements

Using Fig. 48-5 we can establish the *electron configurations* of the various elements and thus tell something about their chemical properties. Consider these three configurations,

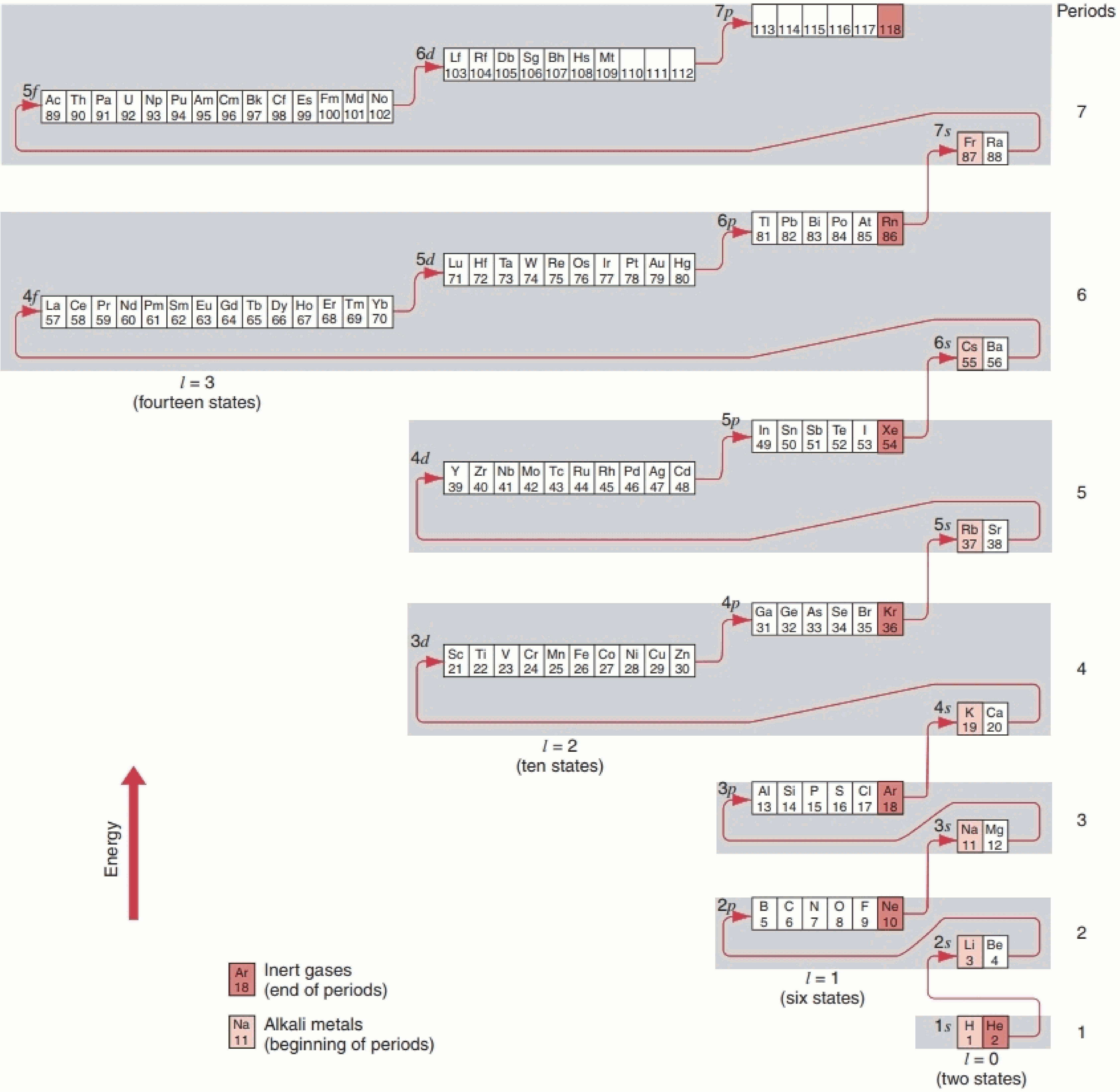

Periods
7p 113 114 115 116 117 118
6d Lr 103 Rf 104 Db 105 Sg 106 Bh 107 Hs 108 Mt 109 110 111 112
5f Ac 89 Th 90 Pa 91 U 92 Np 93 Pu 94 Am 95 Cm 96 Bk 97 Cf 98 Es 99 Fm 100 Md 101 No 102
7s Fr 87 Ra 88
7
6p Tl 81 Pb 82 Bi 83 Po 84 At 85 Rn 86
5d Lu 71 Hf 72 Ta 73 W 74 Re 75 Os 76 Ir 77 Pt 78 Au 79 Hg 80
4f La 57 Ce 58 Pr 59 Nd 60 Pm 61 Sm 62 Eu 63 Gd 64 Tb 65 Dy 66 Ho 67 Er 68 Tm 69 Yb 70
6s Cs 55 Ba 56
6
l = 3
(fourteen states)
5p In 49 Sn 50 Sb 51 Te 52 I 53 Xe 54
4d Y 39 Zr 40 Nb 41 Mo 42 Tc 43 Ru 44 Rh 45 Pd 46 Ag 47 Cd 48
5s Rb 37 Sr 38
5
4p Ga 31 Ge 32 As 33 Se 34 Br 35 Kr 36
3d Sc 21 Ti 22 V 23 Cr 24 Mn 25 Fe 26 Co 27 Ni 28 Cu 29 Zn 30
4s K 19 Ca 20
4
l = 2
(ten states)
3p Al 13 Si 14 P 15 S 16 Cl 17 Ar 18
3s Na 11 Mg 12
3
2p B 5 C 6 N 7 O 8 F 9 Ne 10
2s Li 3 Be 4
2
l = 1
(six states)
1s H 1 He 2
1
l = 0
(two states)
Energy
Ar 18 Inert gases (end of periods)
Na 11 Alkali metals (beginning of periods)

FIGURE 48-5. Starting with hydrogen at the bottom, the curved line passes through each of the seven (shaded) horizontal periods of the periodic table. Each period starts with an alkali metal and ends with an inert gas.

established by starting at the bottom of Fig. 48-5 and working your way upward to the element in question, counting subshells as you go:

Fluorine	$Z = 9$	$1s^2\,2s^2\,2p^5$
Neon	$Z = 10$	$1s^2\,2s^2\,2p^6$
Sodium	$Z = 11$	$1s^2\,2s^2\,2p^6\,3s^1$

In this notation, the first number in each grouping (such as $1s^2$) is the principal quantum number n, which defines a shell; the letter is the code for the angular momentum quantum number, defining the subshell; and the superscript gives the number of electrons occupying that particular subshell. Thus $1s^2$ indicates 2 electrons in the $1s$ subshell. The sum of the superscripts gives the atomic number of the element in question.

As an example, the six electrons in the $2p$ state of neon have $n = 2$ and $l = 1$. Thus, neon has a filled $2p$ subshell, a particularly stable configuration. Neon is correspondingly an inert gas; under most circumstances, it does not form compounds with other elements. The elements directly above neon in Fig. 48-5 and in the periodic table in Appendix E are also inert gases.

Fluorine, on the other hand, has a "hole" in its $2p$ subshell. This subshell can accommodate six electrons but, in fluorine, only five are present. Thus fluorine readily accepts an electron and forms compounds with other elements that are able to provide it.

Sodium has a single electron in a $3s$ subshell, outside an inert neon-like inner core. Sodium readily forms compounds with elements that have a single "hole" available to accept this valence electron. Sodium fluoride (NaF) for example, is a particularly stable compound.

There are occasional small departures from the electron configuration that may be deduced from Fig. 48-5. One such example occurs for copper ($Z = 29$), which we would expect from Fig. 48-5 to have the configuration

$$1s^2\,2s^2\,2p^6\,3s^2\,3p^6\,4s^2\,3d^9.$$

However, it turns out that it is energetically more favorable to complete the filling of the $3d$ subshell (which can hold ten electrons but only has nine in the configuration shown) before putting electrons in the $4s$ subshell. As a result, the actual electron configuration for copper is

$$1s^2\,2s^2\,2p^6\,3s^2\,3p^6\,3d^{10}\,4s^1.$$

This single $4s$ electron is responsible for the large electrical conductivity of copper. There are other examples of such configuration departures.

The energy required to remove the least tightly bound electron from an atom in its ground state is called its *ionization energy.* Figure 48-6 shows the ionization energies of the elements. Note that for each subshell, the ionization energy rises steadily and reaches a maximum at an inert gas, and there is a sharp drop to the alkali element that follows. Sodium, for example, has an ionization energy of 5.14 eV. A much larger energy (47.3 eV) must be supplied to remove a *second* electron from sodium; with one electron removed, the sodium ion has an electron configuration similar to that of neon, which, consisting of three filled subshells, is particularly stable.

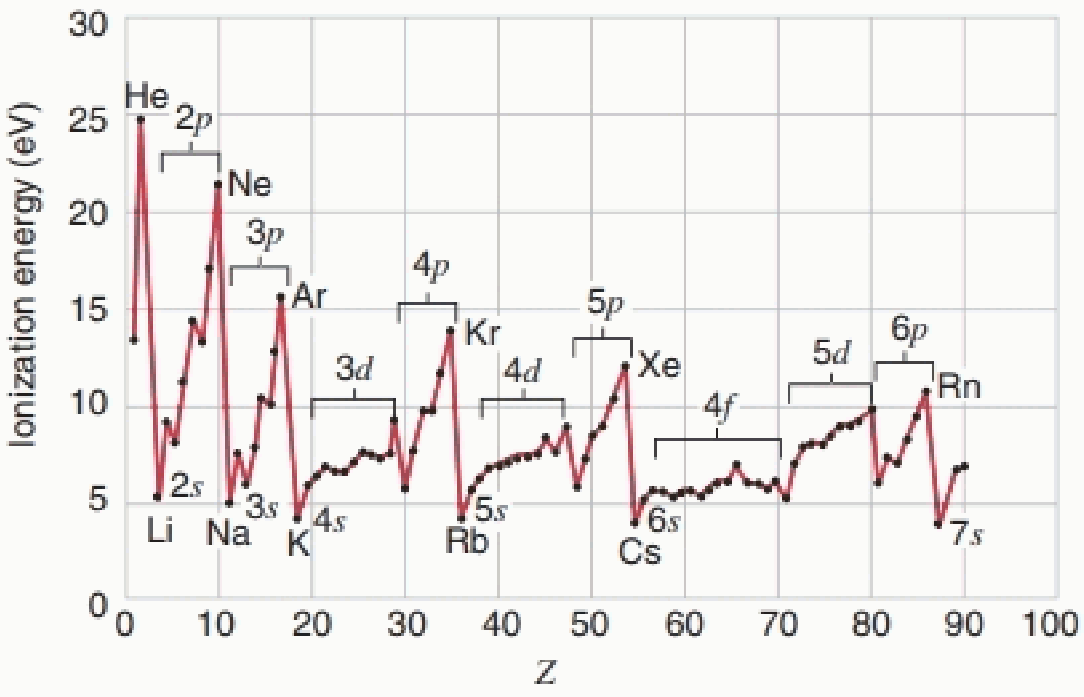


FIGURE 48-6. The ionization energies of the elements, plotted against their atomic number. The subshells of the removable electron are indicated.

Excited States and Optical Transitions

So far we have discussed mostly the minimum-energy or ground-state configurations of atoms. When we add energy to an atom, perhaps by placing it in an electrical discharge tube or bombarding it with radiation, we can raise the atom to one of its higher-energy states.

The energy differences between the ground state and the next-higher energy states to which the outer electron can be excited are typically of the order of a few electron volts. When the electron drops back to its ground state, it emits radiation whose photon energy (which is equal to the difference between the energies of the excited state and the ground state) is in the eV range, corresponding to visible light. For this reason, such changes in the state of the electron are called *optical transitions.*

Figure 48-7 shows some excited states of the single valence electron of sodium and some transitions that may occur. (Not all are in the visible range.) Note that many transitions that are energetically possible simply do not occur. For two examples, there is no transition from the $5s$ to the $4s$ level or from the $4d$ to the $4s$ level. Once we have found the wave functions for the various states, it is possible to calculate the probabilities that there will be transitions between them. When we do so, certain rules, called *selection rules,* emerge from the mathematics. Chief among them in relevance to Fig. 48-7 is that the orbital angular momentum quantum number of the electron making the transition must change by a single unit only; that is,

$$\Delta l = \pm 1. \tag{48-6}$$

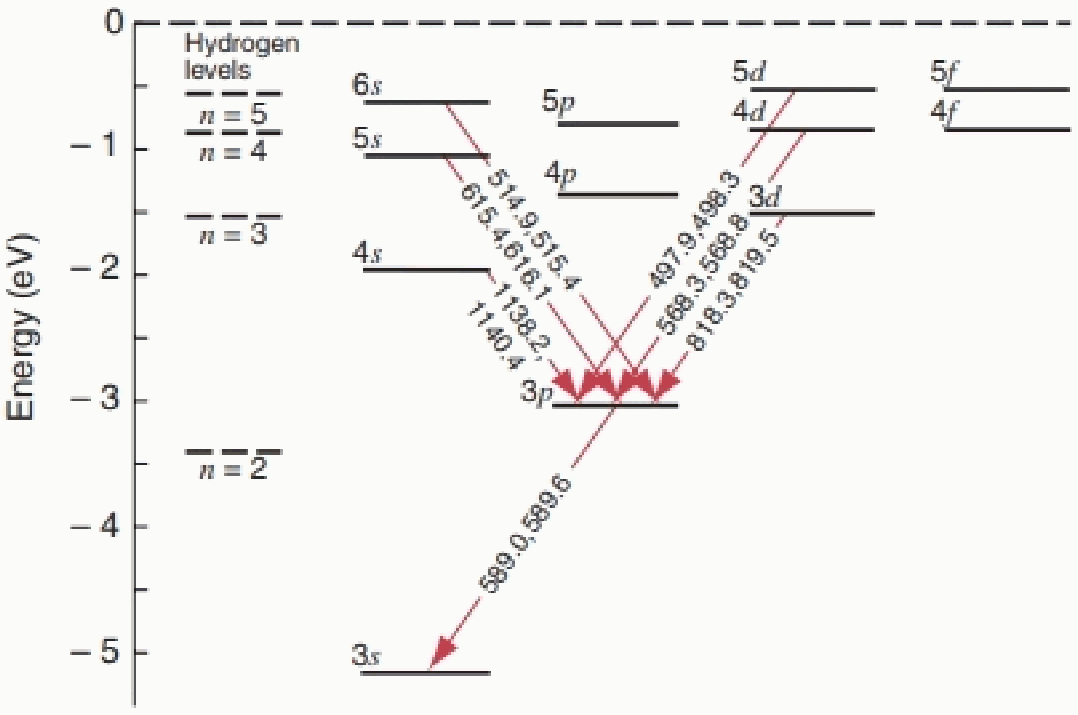


FIGURE 48-7. An energy level diagram for the valence electron of sodium. The levels have been spread out sideways, so that each vertical column contains only levels with the same orbital angular momentum quantum number. The arrows show some transitions that may be emitted, labeled with their wavelengths in nm. (The $3p$ level actually consists of two very closely spaced states, so all transitions involving that level have two components differing only slightly in wavelength; see Section 48-8.)

A transition from the 5*s* to the 4*s* level involves $\Delta l = 0$ and one from the 4*d* to the 4*s* level involves $\Delta l = -2$. Each of these transitions violates the selection rule of Eq. 48-6. Selection rules are not absolute; transitions that violate them may occur but the intensity of the emitted radiation is usually very weak.

48-5 ATOMIC MAGNETISM

So far in this chapter we have discussed what we can learn about the properties of atoms by doing experiments to measure their energy levels, usually by observing the wavelengths or energies of the photons that are emitted when an electron makes a transition from one energy state to another. The angular momentum properties of the atom are most directly studied by placing the atom in a magnetic field. In this section and the next one, we will discuss how the orbital and spin angular momentum influence the atom's magnetic structure. Later we will see that a magnetic field can also affect the wavelengths of the emitted photons.

It is useful to begin this discussion using Bohr's hybrid model, in which we can view the electron in its orbit as a tiny current loop (see Sections 35-1 and 35-3). As we showed in Eq. 35-10, the magnetic dipole moment of such a current loop is (using the present notation)

$$\vec{\boldsymbol{\mu}}_l = -\frac{e}{2m}\vec{\mathbf{L}}, \qquad (48\text{-}7)$$

where $\vec{\boldsymbol{\mu}}_l$ is a vector that represents the magnetic dipole moment and $\vec{\mathbf{L}}$ is the orbital angular momentum vector of the electron. The negative sign comes about because of the negative charge of the electron.

As we discussed in Section 47-6, the uncertainty principle limits our knowledge about $\vec{\mathbf{L}}$ to its magnitude and any one component. It is customary to choose the *z* component as the one that is measured, and so we can write Eq. 48-7 by taking its *z* component as

$$\mu_{lz} = -\frac{e}{2m}L_z. \qquad (48\text{-}8)$$

Although Eq. 48-8 was derived semiclassically, it remains correct in quantum mechanics.

The allowed values of L_z are given by Eq. 47-30 as $m_l h/2\pi$, and so Eq. 48-8 becomes

$$\mu_{lz} = -\frac{e}{2m}\frac{m_l h}{2\pi} = -m_l\frac{eh}{4\pi m}. \qquad (48\text{-}9)$$

We can express the magnetic dipole moment in terms of the *Bohr magneton* μ_B, which is defined as

$$\mu_B = \frac{eh}{4\pi m} = 9.274 \times 10^{-24}\ \text{J/T}. \qquad (48\text{-}10)$$

Just as the Bohr radius is a convenient measure of distance on the atomic scale, so the Bohr magneton is a convenient measure of atomic magnetic moments. Equation 48-9 then becomes

$$\mu_{lz} = -m_l\mu_B, \qquad (48\text{-}11)$$

in which m_l takes the values given by Eq. 47-29.

If we were to place an atom having a magnetic dipole moment $\vec{\boldsymbol{\mu}}_l$ in a magnetic field $\vec{\mathbf{B}}$, which we assume is in the *z* direction, the energy associated with the interaction between the atom and the magnetic field is $U = -\vec{\boldsymbol{\mu}}_l \cdot \vec{\mathbf{B}} = -\mu_{lz}B_z = m_l\mu_B B_z$. That is, atoms with different values of m_l have different energies in the field, which provides a way to determine their orbital angular momentum.

Although we cannot use the semiclassical model to do a similar calculation for the spin angular momentum, a similar result is obtained from a full quantum mechanical calculation. The magnetic dipole moment due to the spin is

$$\mu_{sz} = -g_s m_s \mu_B. \qquad (48\text{-}12)$$

The factor of g_s that appears in Eq. 48-12 does not appear in the equation for the orbital magnetic dipole moment. Like the spin itself, this is purely a quantum mechanical factor that has no classical counterpart. The value of g_s is very close to 2, in contrast to the coefficient in Eq. 48-11, which is equal to 1.

Knowledge of the exact value of g_s is critical, because it can be calculated from the theory that couples quantum mechanics to electromagnetism. The theoretically calculated value is $g_s = 2.0002319305$, and the experimental value is $g_s = 2.000231930437$. In each case the uncertainty is roughly one unit in the last digit. The calculation of this important factor to 11 significant figures and its experimental verification is a triumph of the methods of quantum mechanics and of the experimenters' skills.

The Einstein–de Haas Experiment

An ordinary iron bar shows no external magnetic properties because its elementary atomic magnets are randomly arranged, their effects canceling at all external points. When those elementary magnets are lined up, however, as they are in a bar magnet, their combined magnetic strength is there for all to see; for example, you can pick up small objects such as paper clips.

When the magnetic moments of an assembly of atoms are lined up, their angular momenta—to which they are rigidly coupled—must also be lined up. Because the two quantities are strongly coupled to each other, the magnetism of the atom provides a convenient "handle" by which to explore its angular momentum.

In 1915, Einstein, in collaboration with the Dutch physicist W. J. de Haas, carried out an experiment that explored this relationship between the magnetism of the iron atom and its angular momentum. Bear in mind that when this experiment was carried out, quantum mechanics had not yet been discovered, Bohr's provisional theory of the hydrogen atom was only two years old, and the existence of the spin angular momentum of the electron was still unknown.

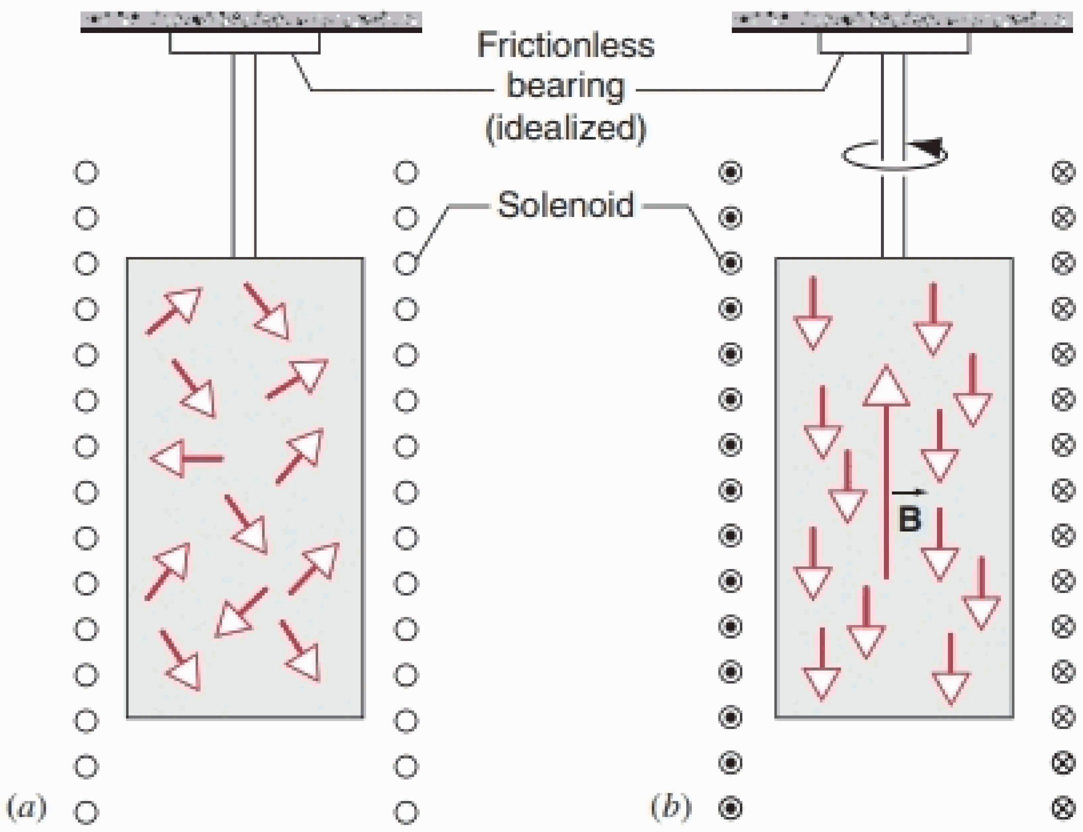


FIGURE 48-8. An idealized adaptation of the experiment of Einstein and de Haas. (*a*) An iron cylinder, free to rotate, is suspended within a solenoid. (*b*) When the magnetic field is turned on, the magnetic moments of the iron atoms tend to line up, as do their angular momentum vectors. To conserve angular momentum, the cylinder must then rotate. The purpose of the experiment is to determine the *g* factor of the iron atoms.

If an iron bar is suddenly magnetized, perhaps by switching on a current in a solenoid as in Fig. 48-8, the individual magnetic moments, and thus also the individual angular momenta, of its atoms suddenly become lined up. Because the angular momentum of the isolated bar must be conserved, the bar as a whole must start to rotate in the opposite sense. This *Einstein–de Haas effect,* as it is called, is small and the measurements are difficult to make.

Knowing nothing of spin, Einstein and de Haas confidently expected the results of their experiment to yield a relationship between magnetic dipole moment and angular momentum similar to Eq. 48-8 or, equivalently, a coefficient of unity in Eq. 48-11. Later experiments, carried out with improved apparatus, instead showed agreement with Eq. 48-12, with $g_s = 1.929 \pm 0.006$. It seems clear that this experiment is measuring the *spin* magnetic dipole moment and not the *orbital* magnetic dipole moment. We now know—as Einstein could not—that the magnetism of the iron atom is associated entirely with the spin magnetic moments of its electrons, the orbital magnetic moments of the individual atomic electrons canceling to zero. Although the numerical results of Einstein and de Haas did not approach the precision of later measurements, their early experiment demonstrated beyond doubt the intimate connection between angular momentum and magnetism.

SAMPLE PROBLEM 48-4. An unmagnetized iron cylinder, of radius $R = 1.0$ mm, hangs from a frictionless bearing so that it can rotate freely about its axis; see Fig. 48-8*a*. A magnetic field is suddenly applied parallel to this axis, causing the magnetic dipole moments of the iron atoms to align themselves parallel to the field. The atomic angular momentum vectors, which are coupled back to back with the magnetic dipole moment vectors, also become aligned and the cylinder will start to rotate in the opposite sense. Find T, the period of rotation of the cylinder. Assume that each iron atom has an angular momentum of $h/2\pi$ and that the iron atoms are completely aligned when the magnetic field is turned on. The molar mass of iron is $M = 0.0558$ kg/mol.

Solution The angular momentum L_{cyl} of the rotating cylinder must be equal in magnitude (though opposite in direction) to L_{atoms}, the angular momentum associated with the aligned atoms. If N is the number of atoms in the cylinder, N_A is the Avogadro constant, and m is the mass of the cylinder, we can write

$$L_{\text{atoms}} = N\left(\frac{h}{2\pi}\right) = \frac{N_A m}{M}\left(\frac{h}{2\pi}\right).$$

For the rotating cylinder we have

$$L_{\text{cyl}} = I\omega = \left(\frac{1}{2}mR^2\right)\left(\frac{2\pi}{T}\right),$$

in which I is the rotational inertia of the cylinder about its symmetry axis and ω is its angular speed. Equating these two expressions and solving for T yields

$$\begin{aligned} T &= \frac{(2\pi^2)R^2M}{N_A h} \\ &= \frac{(2\pi^2)(1 \times 10^{-3}\ \text{m})^2(0.0558\ \text{kg/mol})}{(6.02 \times 10^{23}\ \text{mol}^{-1})(6.63 \times 10^{-34}\ \text{J}\cdot\text{s})} \\ &= 2760\ \text{s} = 46\ \text{min}. \end{aligned}$$

Actually, Einstein and de Haas suspended their cylinder from a torsion fiber and used more refined observing techniques than we have described here.

48-6 THE STERN–GERLACH EXPERIMENT

The angular momentum vector associated with an atomic state can take up only certain specified directions in space with respect to an arbitrarily chosen axis; see Fig. 47-13. This concept of *space quantization* was predicted theoretically by Wolfgang Pauli. In 1922, Otto Stern (German, 1888–1969) and Walther Gerlach (German, 1889–1979) sought to verify this prediction experimentally.

Figure 48-9 shows the apparatus of Stern and Gerlach. Silver is vaporized in an electrically heated "oven," and silver atoms spray into the evacuated apparatus through a small hole in the oven wall. The atoms (which are electrically neutral but have a magnetic moment) are formed into a narrow beam as they pass through a slit in a screen. The beam, thus collimated, then passes between the poles of an electromagnet and, finally, deposits its silver atoms on a glass plate that serves as a detector.

The pole faces of the magnet are shaped to make the magnetic field as *nonuniform* as possible. The beam of silver atoms passes very close to the sharp V-shaped ridge in

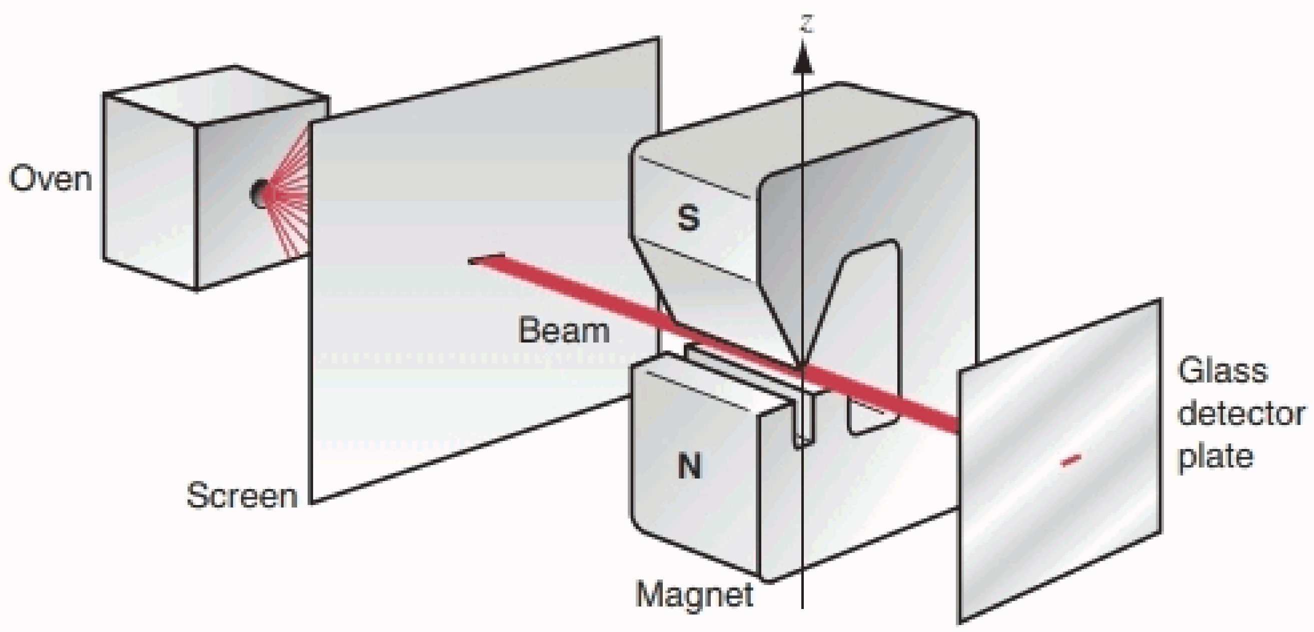


FIGURE 48-9. The apparatus of Stern and Gerlach. A beam of silver atoms passes through a nonuniform magnetic field and is deposited on a glass plate. Because the atoms possess a magnetic moment, they will be deflected by the magnetic field. The purpose of the experiment is to demonstrate the reality of space quantization.

the upper pole face, where the nonuniformity of the field is greatest.

A Dipole in a Nonuniform Field

Figure 48-10*a* shows a dipole of magnetic moment $\vec{\mu}$ at an angle θ with a *uniform* magnetic field. We can represent the dipole as a tiny bar magnet, with the magnetic dipole moment vector $\vec{\mu}$ pointing (by convention) from its south pole to its north pole. We consider the forces exerted by the magnetic field on the magnet to be concentrated at its poles, as shown in the figure. We see that, for a uniform field, a net torque acts on the magnet (for $\theta \neq 0$ or 180°); however, there is no net force on the dipole, no matter what the value of θ, because the upward and downward forces on the poles are of the same magnitude and they cancel.

Figures 48-10*b* and *c* show the situation in a *nonuniform* field. Here the upward and downward forces do not have the same magnitude because the two poles are immersed in fields of different strengths. In this case there *is* a net force, its magnitude and direction depending on the orientation of the dipole—that is, on the value of θ. In Fig. 48-10*b* the net force is up and in Fig. 48-10*c* it is down. Thus the silver atoms in the beam of Fig. 48-9, as they pass through the electromagnet, are deflected up or down, depending on the orientation of their magnetic dipole moments with respect to the z direction.

Let us calculate the deflecting force. The potential energy of a magnetic dipole in a magnetic field $\vec{\mathbf{B}}$ is given by Eq. 35-6,

$$U = -\vec{\mu}\cdot\vec{\mathbf{B}}.$$

From symmetry (see Fig. 48-10*b*, *c*), the magnetic field at the beam position has no x or y components. Thus $B = B_z$, and we can write the potential energy as

$$U = -\mu_z B_z. \tag{48-13}$$

The net force F_z on the dipole is $-dU/dz$ or,

$$F_z = \mu_z \frac{dB_z}{dz}. \tag{48-14}$$

Thus the net force depends, not on the magnitude of the field itself, but on its *spatial derivative,* or *gradient.* In Figs. 48-10*b* and *c*, B_z increases as z increases, so the derivative

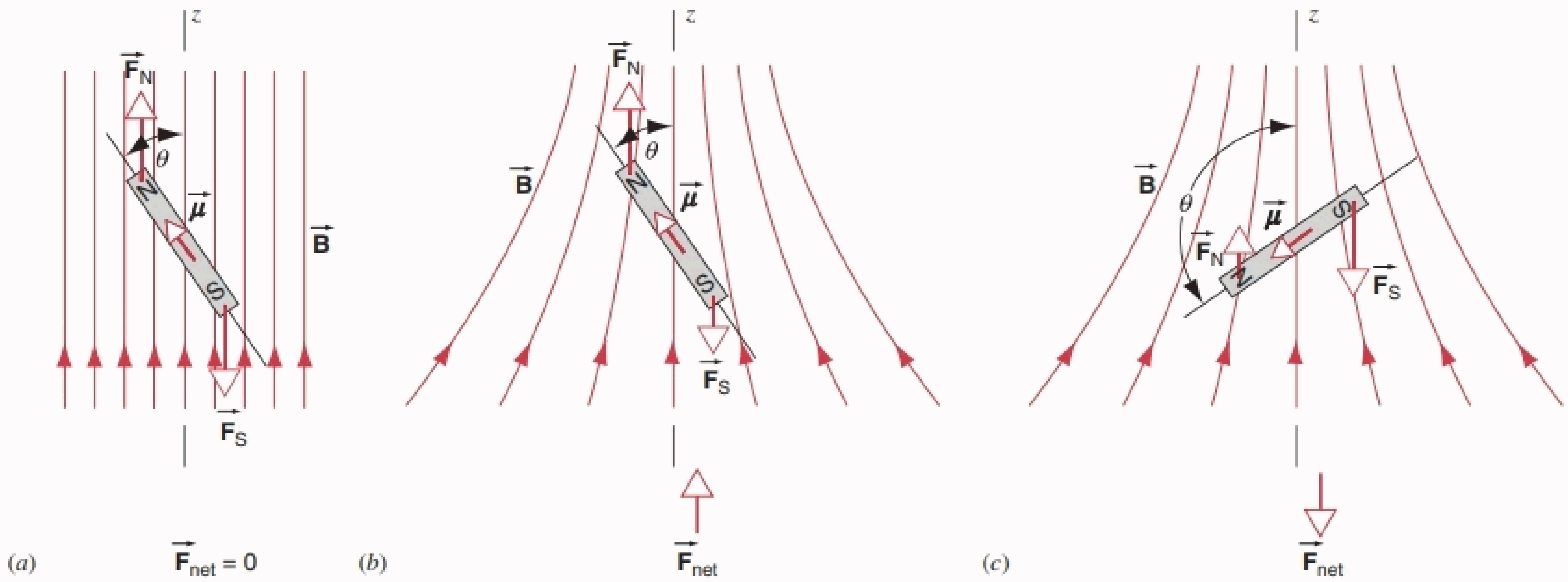


FIGURE 48-10. (*a*) A magnetic dipole, represented as a small bar magnet, is placed in a uniform magnetic field. No net force acts on it. (*b*) In a nonuniform field, for the orientation shown, there is a net force in the z direction. (*c*) For this orientation, there is a net force in the negative z direction.

dB_z/dz is positive. Thus the sign of the deflecting force F_z in Eq. 48-14 depends on the sign of μ_z. If μ_z is positive (as in Fig. 48-10*b*) the silver atom is deflected upward; if it is negative (as in Fig. 48-10*c*) the deflection is downward.

The Experimental Results

What should Stern and Gerlach have expected to see when they examined the silver deposited on the glass detector plate? If space quantization did not exist, then μ_z in Eq. 48-13 could take on *any* value from $+\mu$ to $-\mu$; the result would be a spreading out of the beam when the magnet was turned ON. What Stern and Gerlach *did* see, however, was that the beam was split cleanly into two subbeams, each subbeam corresponding to one of the two permitted orientations of the magnetic moment of the silver atom.

Figure 48-11, a graph of beam intensity as a function of detector position, shows the results of a recent repetition of the Stern–Gerlach experiment. These experimenters used cesium atoms rather than silver atoms and employed a more sensitive detector but, otherwise, the arrangement was essentially unchanged. The separation of the beam into two subbeams when the magnetic field is turned from OFF to ON is especially clear.

We know now (but it was not known then) that all the spin and orbital magnetic moments of the electrons in a silver (or a cesium) atom cancel, except for those of the atom's single valence electron. For this electron, the orbital magnetic moment is zero (because the orbital angular momentum is zero), leaving only the spin magnetic moment. As Eq. 47-38 shows, this can take up only two orientations in a magnetic field, corresponding to $m_s = +1/2$ and $m_s = -1/2$. Hence there are *two* subbeams—and not some other number. Stern and Gerlach ended the published report of their work with the words: "We view these results as direct experimental evidence of space quantization in a magnetic field." Physicists everywhere agreed. For these experiments and others using beams of atoms, Stern was awarded the 1943 Nobel Prize in physics.

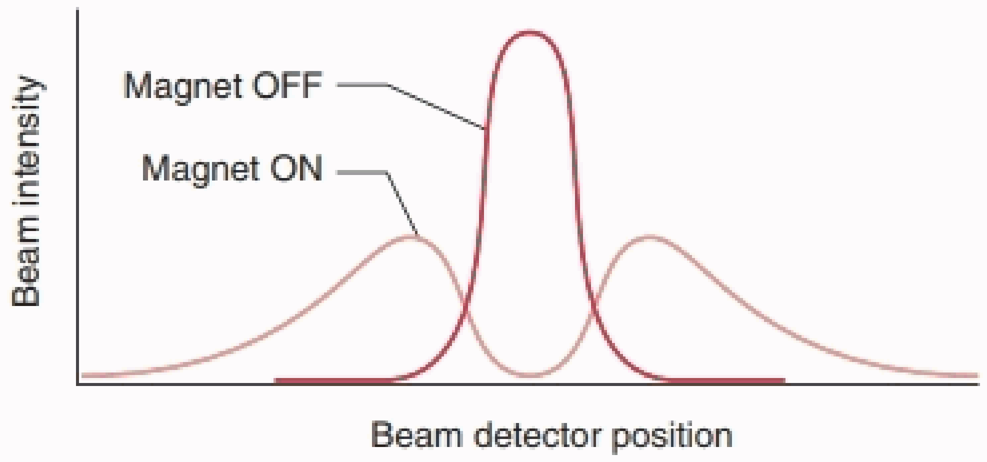


FIGURE 48-11. The central peak shows that, in a modern repetition of the Stern–Gerlach experiment, the atomic beam is not deflected when the magnet is turned off (or operated at very low power). The double-peak curve shows that the atoms in the beam are either aligned or anti-aligned with the magnetic field; other orientations do not occur.

SAMPLE PROBLEM 48-5. The magnet in a Stern–Gerlach experiment has a magnetic field gradient $dB_z/dz = 1.4$ T/mm at the location of the beam of silver atoms. The length D of the beam path through the magnet is 3.5 cm. The temperature of the oven in which the silver is evaporated is adjusted so that the most probable speed v for the atoms in the beam is 750 m/s. Find the vertical deflection of either subbeam as it emerges from the magnet. (The mass m of a silver atom is 1.7×10^{-25} kg and its effective magnetic moment is 9.27×10^{-24} J/T.)

Solution The vertical acceleration of a silver atom as it passes through the magnet is, from Newton's second law and Eq. 48-14,

$$a = \frac{F_z}{m} = \frac{\mu_z (dB_z/dz)}{m}.$$

Moving horizontally at speed v, each silver atom passes through the length D of the magnet in a time $t = D/v$. The vertical deflection of the atom as it clears the magnet is then

$$d = \tfrac{1}{2}at^2 = \frac{\mu_z(dB_z/dz)D^2}{2mv^2}.$$

Substituting numerical values yields

$$d = \frac{(9.27 \times 10^{-24}\ \mathrm{J/T})(1.4 \times 10^3\ \mathrm{T/m})(3.5 \times 10^{-2}\ \mathrm{m})^2}{(2)(1.7 \times 10^{-25}\ \mathrm{kg})(750\ \mathrm{m/s})^2}$$

$$= 8.31 \times 10^{-5}\ \mathrm{m} \cong 0.083\ \mathrm{mm}.$$

The separation between the two subbeams is twice this, or about 0.16 mm. This separation is not large, but it is readily measurable.

48-7 NUCLEAR MAGNETIC RESONANCE

The proton, like the electron, has a spin of 1/2 and thus the z component of its associated magnetic moment can have two values only, corresponding to $m_s = +1/2$ and $m_s = -1/2$. The fact that many particles, such as the proton, have spin has proved to be of great practical value through the phenomenon of *nuclear magnetic resonance,* (NMR) discovered independently by 1946 by Felix Bloch and Edward Purcell. These two investigators shared the 1952 Nobel Prize for their work.

When a proton is placed in a magnetic field $\vec{\mathbf{B}}$, an energy change of $2\mu_z B_z$ occurs when its spin magnetic moment "flips" between an orientation parallel to the field and another antiparallel to the field (Fig. 48-12*a*). Such a spin flip can be triggered by subjecting the protons to a relatively weak "tickler" alternating magnetic field whose direction is at right angles to $\vec{\mathbf{B}}$ and whose frequency f is selected such that

$$hf = 2\mu_z B_z. \tag{48-15}$$

As Fig. 48-12*b* suggests, the "spin flip" may be from a state with lower energy to one of higher energy or in the reverse direction. In the first case there is a *decrease* in the energy of the alternating magnetic field and in the second case there is

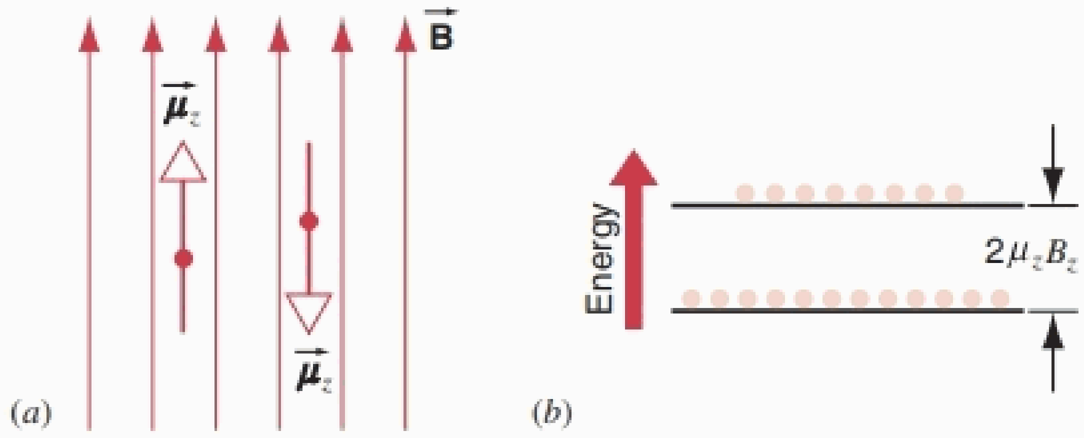


FIGURE 48-12. (*a*) A proton, whose spin is 1/2, can take only two orientations in an applied magnetic field. If Eq. 48-15 is satisfied, the protons can be induced to flip from one orientation to the other. (*b*) Normally, there are more protons in the lower energy state than in the higher energy state, so the result is a net absorption of energy by the entire proton-containing sample.

an *increase.* As the figure suggests, however, in thermal equilibrium there are normally more protons in the lower state than in the upper. Thus the net effect is a (detectable) absorption of energy from the alternating magnetic field.

The field $\vec{\mathbf{B}}$ consists of an external field $\vec{\mathbf{B}}_{\text{ext}}$ (perhaps due to an electromagnet) and an internal magnetic field $\vec{\mathbf{B}}_{\text{int}}$ (due to the chemical environment in which the proton finds itself). For example, in a molecule of ethanol, whose formula we may write as $CH_3–CH_2–OH$, each hydrogen nucleus experiences a different internal field because of its location within the molecule. The resonance condition for a spin flip is then

$$2\mu_z(B_{\text{ext}} + B_{\text{int}}) = hf, \tag{48-16}$$

in which we assume that $\vec{\mathbf{B}}_{\text{ext}}$ and $\vec{\mathbf{B}}_{\text{int}}$ are in the same direction (the z direction). By keeping the frequency f fixed and varying the external field B_{ext}, we can find several instances at which spin flips occur, each corresponding to a particular environment of a hydrogen nucleus (proton) in an ethanol molecule; see Fig. 48-13.

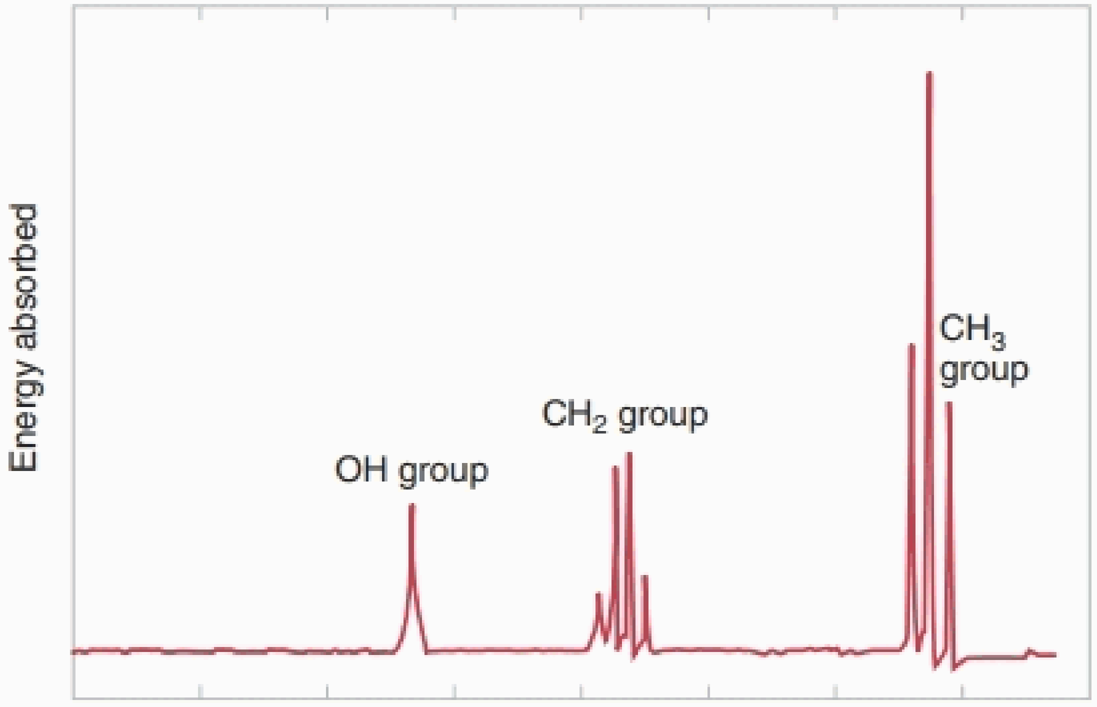


FIGURE 48-13. A nuclear magnetic resonance spectrum of ethanol. All the lines are due to the absorption of the incident radiations when the proton spin flips. The groups of lines correspond to different groupings of hydrogen atoms within the molecule. The entire range of the horizontal axis, which represents the external magnetic field B_{ext}, is considerably less than 10^{-4} T.

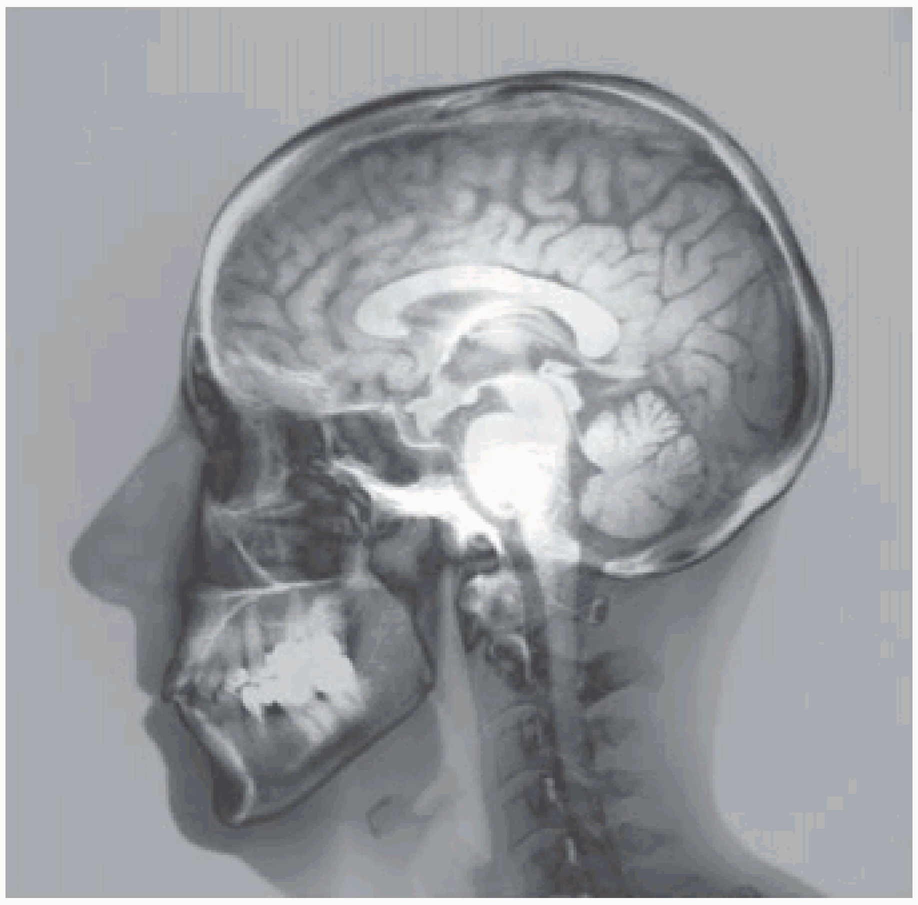


FIGURE 48-14. An MRI image of the cross section of a human head.

Nuclear magnetic resonance has many applications in chemistry and physics. It is also the basis for the highly successful and familiar diagnostic tool called *magnetic resonance imaging* (MRI); see Fig. 48-14.

SAMPLE PROBLEM 48-6. A drop of water is suspended in a magnetic field $\vec{\mathbf{B}}$ whose magnitude is 1.80 T. An alternating "tickler" magnetic field is also applied, its frequency adjusted to generate spin flips of the protons in the water sample. The z component of the proton magnetic moment is $\mu_z = 1.41 \times 10^{-26}$ J/T. What are the frequency f and the wavelength λ of the alternating magnetic field? Assume that local internal magnetic fields are negligible compared to $\vec{\mathbf{B}}$.

Solution From Eq. 48-16 we have, assuming that $B_{\text{int}} = 0$,

$$f = \frac{2\mu_z B}{h} = \frac{(2)(1.41 \times 10^{-26}\ \text{J/T})(1.80\ \text{T})}{6.63 \times 10^{-34}\ \text{J}\cdot\text{s}}$$

$$= 7.66 \times 10^7\ \text{Hz} = 76.6\ \text{MHz}.$$

The corresponding wavelength is

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^8\ \text{m/s}}{7.66 \times 10^7\ \text{Hz}} = 3.92\ \text{m}.$$

This frequency and wavelength are in the short radio-wave band of the electromagnetic spectrum.

48-8 MAGNETISM AND ATOMIC RADIATIONS (Optional)

When we inspect the spectral lines of an atom under high resolution, we often find that what appears to be a single line is actually a pair of very closely spaced lines. We will examine this *fine structure,* as it is called, for the spectrum of the sodium atom ($Z = 11$).

The four quantum numbers that describe the quantum states of the single electron in the hydrogen atom (see Table 47-2) can still be used to describe the quantum states of individual electrons in atoms with more than one electron. We have seen in Section 48-4 that the neutral sodium atom has a single valence electron outside a neon-like core formed by the remaining ten electrons. The angular momenta and the magnetic moments of the electrons in a filled subshell add vectorially to zero so that these properties of the atom depend only on the properties of this single valence electron.

Figure 48-15 is a (partial) energy-level diagram for the sodium atom's valence electron (see also Fig. 48-7). The lowest level, the ground state, is a 3*s* state, for which $n = 3$ and $l = 0$. The upper pair of levels represents a 3*p* state, for which for $n = 3$ and $l = 1$. We see here that, for multielectron atoms, it is no longer true that the energy of a quantum state depends only on the single quantum number n; it also depends on the orbital quantum number l. This is perhaps understandable because the potential energy of the sodium valence electron is associated not only with the nuclear charge, but also with the charges of the other ten electrons; the extent to which this valence electron will penetrate this "electron cloud" depends on the value of l.

The splitting of the 3*p* state into two components is the origin of the fine structure of the sodium atom spectrum. The light emitted when an electron in either of these two 3*p*-level components makes a transition to the ground state forms the so-called *sodium doublet.* This is the intense yellow light emitted, for example, by sodium street lamps. The figure is not to scale; the actual splitting of the 3*p* level is only about 1/1000 of the energy difference between the 3*p* and the 3*s* levels.

Explaining the Fine Structure

The splitting of the 3*p* state comes about because the (z component of the) magnetic moment associated with the spin of the electron may either be parallel or antiparallel to the (z component of the) magnetic moment associated with the orbital moment of the electron. You can think of these two internal atomic magnets as two tiny bar magnets aligned side by side with their magnetic moment vectors either parallel or antiparallel. These two configurations differ in energy because it takes work to change from the antiparallel to the parallel configuration.

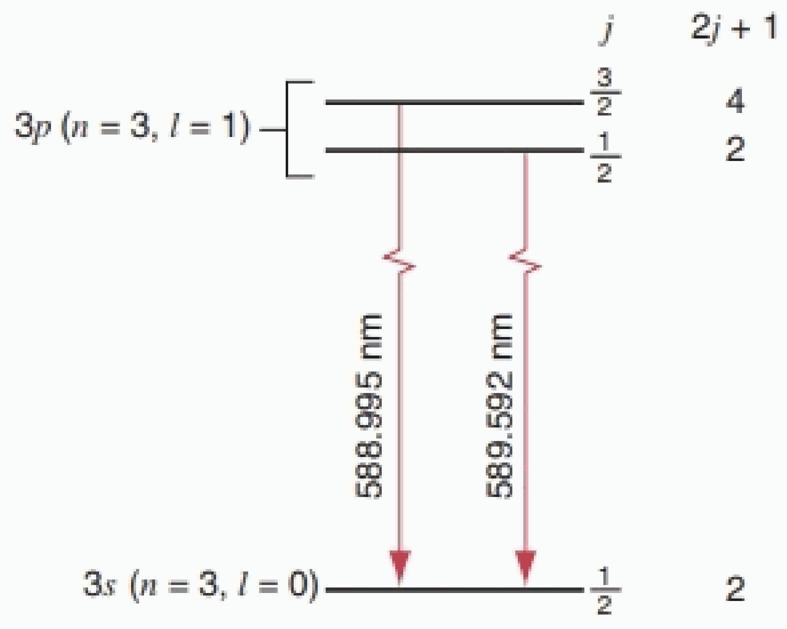


FIGURE 48-15. The 3*s* and the 3*p* levels for the valence electron of sodium. The 3*p* level is split because of the interaction between the orbital and the spin magnetic moments of the valence electron. The total angular momentum quantum number j is shown on the right.

Formally, we describe the situation by introducing a new quantum number j, called the *total angular momentum quantum number,* and defined from

$$j = l \pm \frac{1}{2}. \tag{48-17}$$

This number has the significance that the z component of the total angular momentum vector (including both spin and orbital motion) has the value $m_j(h/2\pi)$, where m_j ranges in value from $-j$ to $+j$, in integral steps.

For the 3*s* state of Fig. 48-15, for which $l = 0$ and $j = 1/2$, there is no orbital magnet moment for the spin magnetic moment to interact with and thus no fine-structure splitting.

The 3*p* state, however, for which $l = 1$, is split into two components, corresponding to its two allowed values of j; namely (see Fig. 48-15),

$$j = 1 + 1/2 = 3/2 \quad \text{and} \quad j = 1 - 1/2 = 1/2$$

The state with $j = 3/2$ has 4 substates, corresponding to $m_j = +3/2, +1/2, -1/2$, and $-3/2$. The state with $j = 1/2$ has 2 substates, corresponding to $m_j = +1/2$ and $-1/2$. The two split states thus have $4 + 2$ or 6 substates between them. In the absence of level splitting, the 3*p* state would have $2(2l + 1) = 2(2 \times 1 + 1) = 6$ substates. Thus fine-structure splitting does not change the number of states; it simply redistributes them.

The Zeeman Effect

Michael Faraday (1791–1867), known for his remarkable intuition, had an idea that the spectrum of a light source would change if you put the source in a strong magnetic field. However, he was not successful in carrying out this experiment because the equipment of his day was not sensitive enough. About 30 years later the Dutch physicist Pieter Zeeman (1865–1943) repeated the experiment with more sensitive apparatus and observed that the spectral lines emitted by a light source were measurably broadened when the source was placed in a strong magnetic field. For this work Zeeman shared the 1902 Nobel Prize. With modern equipment it is readily possible to see the single spectrum lines dividing into discrete components.

Figure 48-16 shows an example of this *Zeeman effect,* as it is called, for light emitted by atoms of rhodium. It is a triumph of quantum mechanics that the number of components into which a given spectrum line is split, their separations, intensities, and polarizations, can be predicted and are in full agreement with experiment.

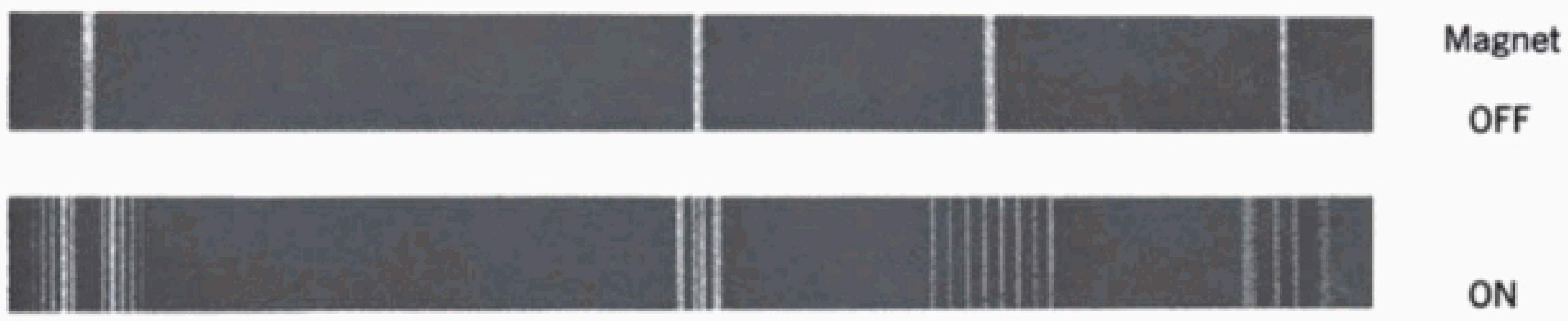


FIGURE 48-16. The Zeeman effect in rhodium. The bottom spectrum shows the splitting of the spectrum lines when the magnetic field is turned on.

To understand the origin of the Zeeman effect, consider the energy-level diagram of Fig. 48-17*a*, which shows (compare Fig. 48-15) only the $j = 3/2$ excited state and the $j = 1/2$ ground state of the valence electron of the sodium atom. Now let the atom be immersed in a strong, uniform external magnet field, whose direction we take as a z axis. We have seen that the $j = 3/2$ state contains four substates, corresponding to four different orientations of the total angular momentum vector with respect to the z axis—that is, to the four allowed values of m_j. These substates will have different energies in the external magnetic field so that the $j = 3/2$ level is split into four components. Similarly, the $j = 1/2$ ground state is split into two components. As the vertical arrows show, when the external magnet is turned ON, the wavelength of one member of the sodium doublet shown in Fig. 48-15 ($\lambda = 588.995$ nm) is split into six closely spaced spectrum lines. (The other member of the doublet, shown in Fig. 48-15 but not in Fig. 48-17, is split into four closely spaced lines.)

By studying Fig. 48-17*b* carefully, you might expect to find eight Zeeman components, not the six that are actually observed. However, two of these transitions are forbidden by the rules of quantum mechanics, one of which permits m_j to change by only zero or ± 1. Can you verify that the six observed lines are consistent with this *selection rule* and that the two "forbidden" lines correspond to $\Delta m_j = \pm 2$, in violation of the rule?

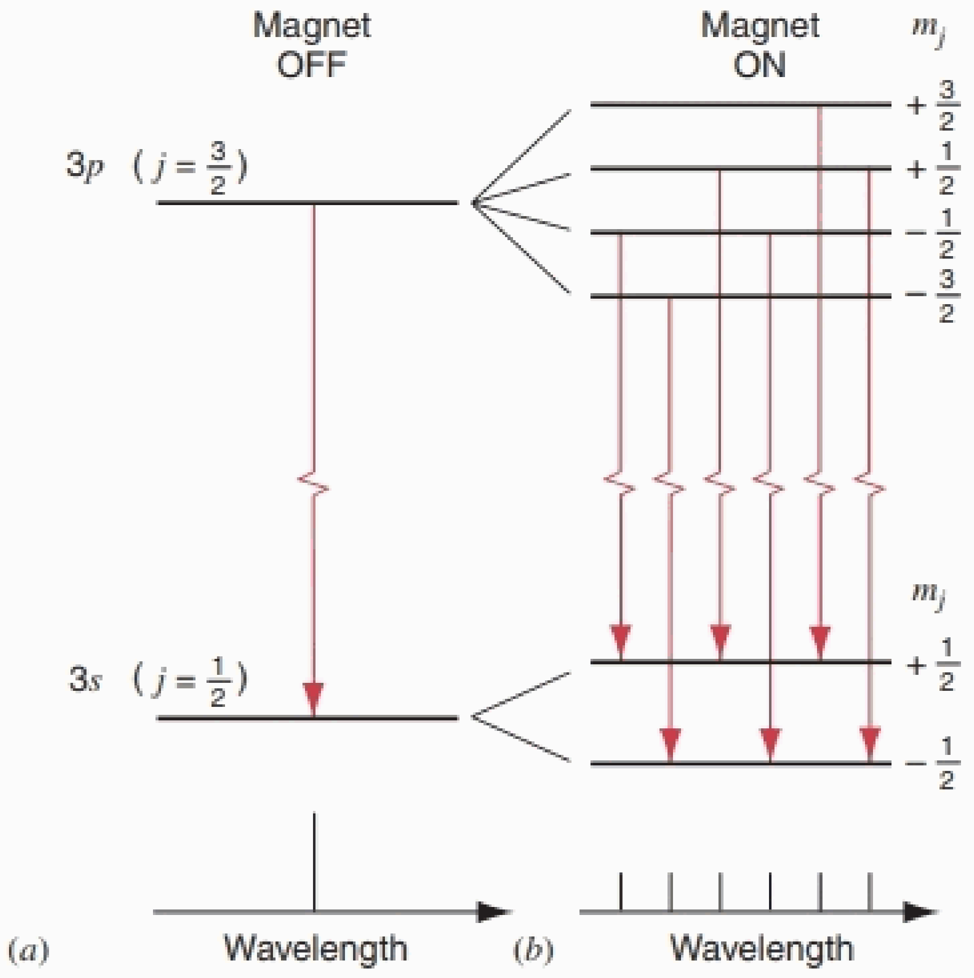


FIGURE 48-17. An energy level diagram illustrating the Zeeman splitting of one component of the sodium doublet; compare closely with Fig. 48-7. When the magnetic field is turned on, the upper level splits into four components, the lower level into two.

48-9 LASERS AND LASER LIGHT

In the late 1940s and again in the early 1960s, quantum mechanics made two enormous contributions to technology—the transistor and the laser. The first stimulated the growth of *electronics,* which deals—at the quantum level—with the interaction between electrons and bulk matter. The laser has lead to the development of *photonics,* which deals with the interaction—again at the quantum level—between photons and bulk matter.

Let us examine some characteristics of laser light (see Fig. 48-18). We shall compare it as we go along with the light emitted by such sources as a tungsten filament lamp (continuous spectrum) or a neon gas discharge tube (line spectrum). We shall see that referring to laser light as "the light fantastic" goes far beyond whimsy.

1. *Laser light is highly monochromatic.* Tungsten light, spread over a broad spectrum, gives us no basis for comparison. The light from select lines in a gas discharge tube can have wavelengths in the visible range that are precise to about 1 part in 10^6. The sharpness of definition of laser

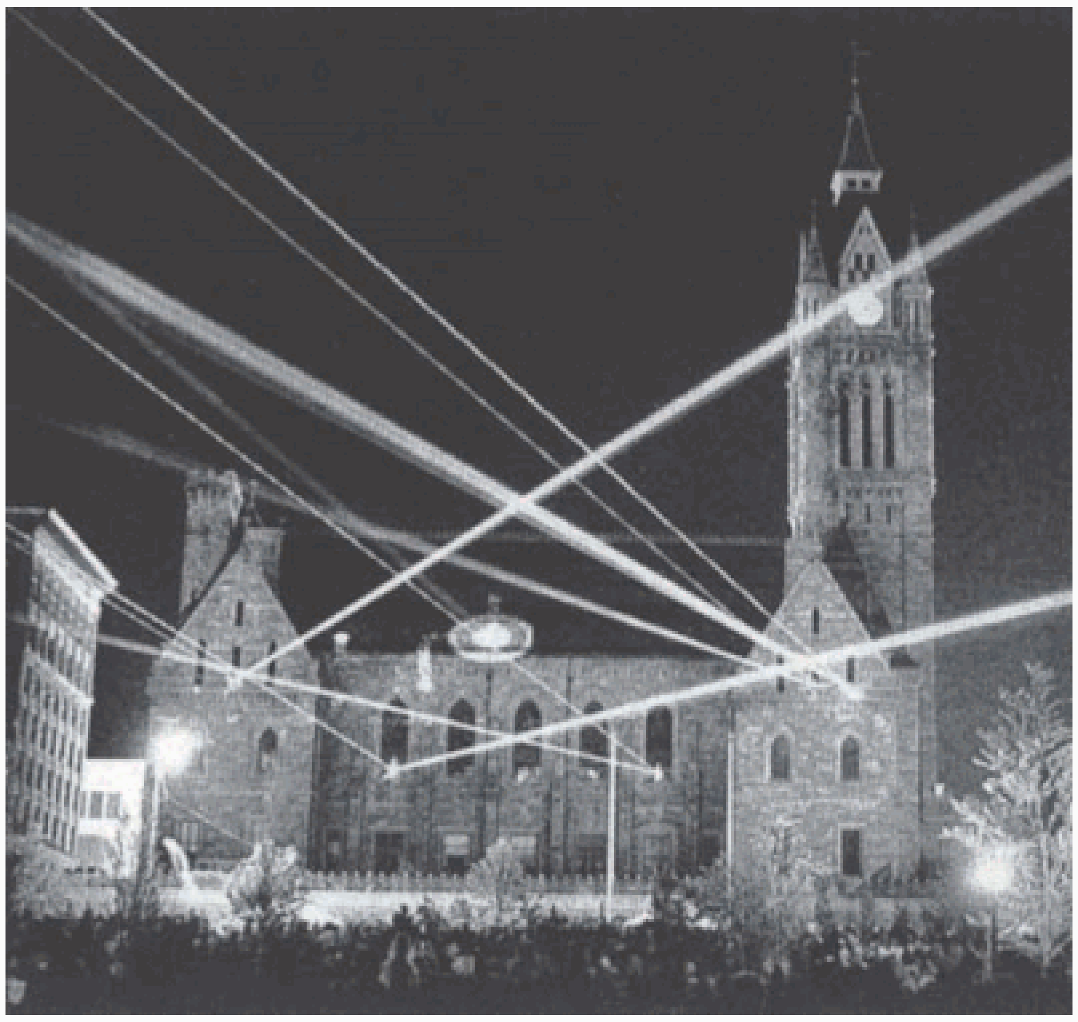


FIGURE 48-18. Laser beams light up the sky!

light can easily be at least a thousand times greater, or 1 part in 10^9.

2. *Laser light is highly coherent.* Wave trains for laser light may be several hundred kilometers long. Interference fringes can be set up by combining two beams that have followed separate paths whose lengths differ by as much as this amount. The corresponding coherence length for light from a tungsten lamp or a gas discharge tube is of the order of a few meters.

3. *Laser light is highly directional.* A laser beam departs from strict parallelism only because of diffraction effects determined (see Section 42-4) by the wavelength of the light and the dimensions of the exit aperture. Light from other sources can be formed into an approximately parallel beam by a lens or a mirror. However, the angular divergence of a beam from a tungsten filament lamp is determined, not so much by diffraction, as by the dimensions of the filament.

4. *Laser light can be sharply focused.* Flux densities for focused laser light of 10^{15} W/cm^2 are readily achieved. By comparison, an oxyacetylene flame has a flux density of only 10^3 W/cm^2.

The smallest lasers, used in CD players and for telephone communication over optical fibers, have as their active medium a semiconducting gallium arsenide crystal about the size of a pin head. The largest lasers, used for laser fusion research, fill a large building. Laser light is used for spot-welding detached retinas, drilling tiny holes in diamonds for drawing fine wires, cutting cloth (50 layers at a time, with no frayed edges) in the garment industry, precision surveying, precision length measurements by interferometry, and the generation of holograms. Laser beams have been reflected from the Moon and from several of the planets.

Einstein and the Laser

In 1917, Einstein, in the process of providing an especially simple and direct proof of Planck's radiation equation (Eq. 45-6) introduced into physics a new concept—that of *stimulated emission.* As it happens, this process is basic to laser operation. Even though the first operating laser did not appear until 1960, the groundwork for its invention was put in place by Einstein's work. The importance of stimulated emission is indicated by the fact that the very word "laser" is an acronym for "*L*ight *A*mplification by the *S*timulated *E*mission of *R*adiation."

It is interesting that, in 1905, Einstein first advanced the concept of the photon and realized that it could provide an explanation for the photoelectric effect. We learn from both of these examples that practical devices of major importance can flow from the study of problems that seem to have no relevance to technology. When next you see a photoelectric elevator door opener or listen to a CD player, think of Einstein.

Now let us examine three processes that involve the interaction between matter and radiation.

1. *Absorption.* Figure 48-19*a* suggests an atomic system in the lower of two possible states, of energies E_1 and E_2. A continuous spectrum of radiation is present. Let a photon from this radiation field approach the two-level atom and interact with it, and let the associated frequency f of the photon be such that

$$hf = E_2 - E_1. \tag{48-18}$$

The result is that the photon vanishes and the atomic system moves to its upper energy state. We call this process *absorption.*

2. *Spontaneous emission.* In Fig. 48-19*b* the atomic system is in its upper state and there is no radiation nearby. After a mean time τ, this (isolated) atomic system moves of its own accord to the state of lower energy, emitting a photon of energy hf ($= E_2 - E_1$) in the process. We call this process *spontaneous emission,* in that no outside influence triggered the emission.

Normally the mean life τ for spontaneous emission by excited atoms is of the order of 10^{-8} s. However, there are some states for which τ is much longer, perhaps 10^{-3} s. We call such states *metastable;* they play an essential role in laser operation. (They have such long lifetimes because they can emit radiation only through processes that violate the selection rule of Eq. 48-6.)

The light from a glowing lamp filament is generated by spontaneous emission. Photons produced in this way are totally independent of each other. In particular, they have different directions and phases. Put another way, the light they produce has a low degree of coherence.

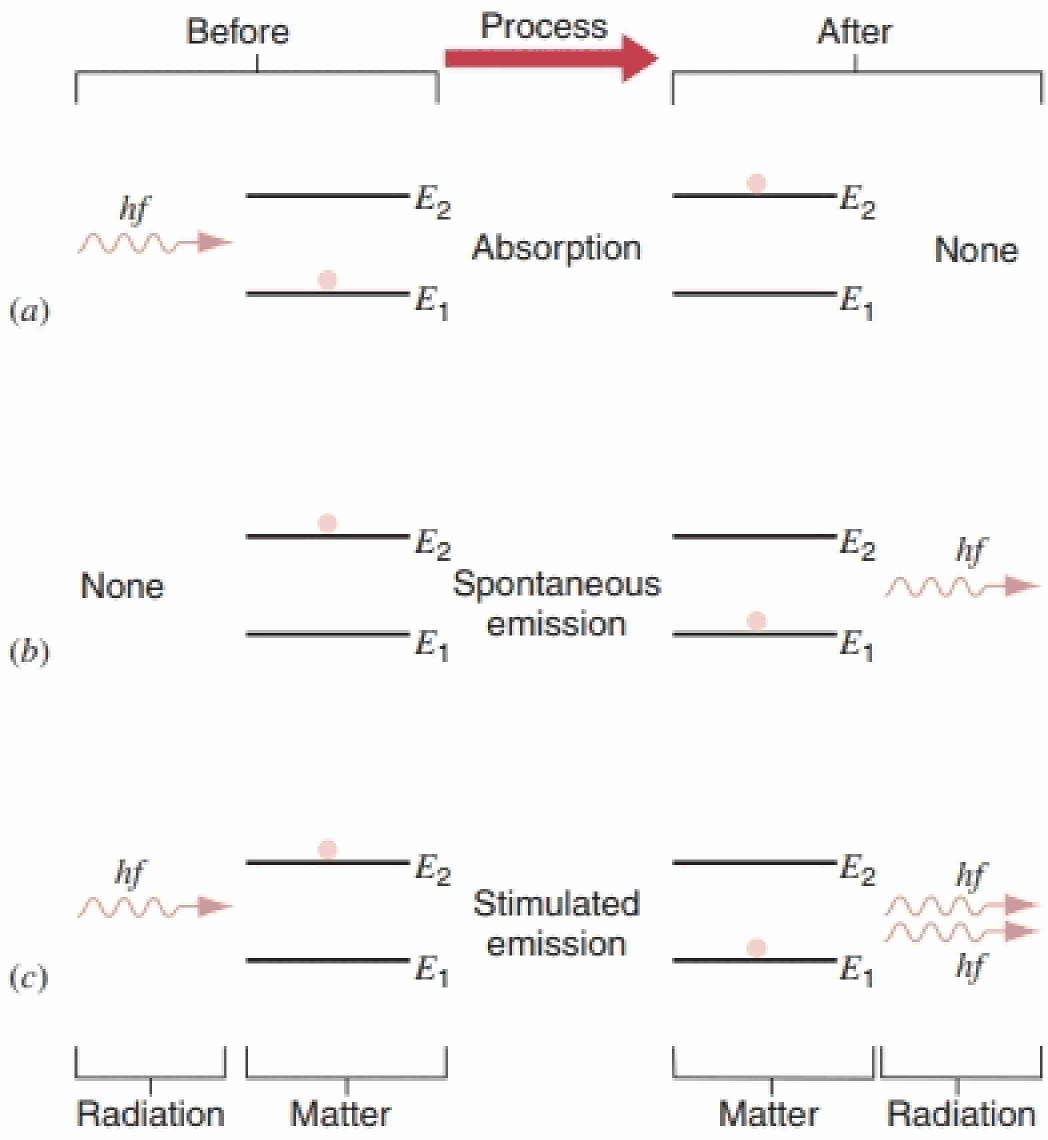


FIGURE 48-19. The interaction of matter and radiation for the processes of (*a*) absorption, (*b*) spontaneous emission, and (*c*) stimulated emission.

3. *Stimulated emission.* In Fig. 48-19*c* the atomic system is again in its upper state, but this time radiation of frequency given by Eq. 48-18 is present. As in absorption, a photon of energy *hf* interacts with the system. The result is that the system is driven down to its lower state, and there are now *two* photons where only one existed before. We call this process *stimulated emission.*

The emitted photon in Fig. 48-19*c* is in every way identical with the "triggering" or "stimulating" photon. It has the same energy, direction, phase, and state of polarization. Furthermore, each of these two photons can cause another stimulated emission event, giving a total of four photons, which can cause additional stimulated emissions, and so on. We can see how a chain reaction of similar processes could be triggered by one such event. This is the "amplification" of the laser acronym. The photons have identical energies, directions, phases, and states of polarization. This is how laser light acquires its characteristics.

Figure 48-19 refers to the interaction with radiation of a single atom. In the usual case, however, we find ourselves dealing with a large number of atoms. For the two-level system of Fig. 48-19, how many of these atoms will be in level E_1 and how many in level E_2? In any system at thermal equilibrium, the number occupying a state at energy E is determined by the exponential Boltzmann factor $e^{-E/kT}$ (see Section 22-5). The ratio of the number of atoms in the upper level to the number in the lower level is

$$n(E_2)/n(E_1) = e^{-(E_2-E_1)/kT}. \qquad (48\text{-}19)$$

Figure 48-20*a* illustrates this situation. The quantity kT is the mean energy of agitation of an atom at temperature T, and we see that the higher the temperature, the more atoms—on long-term average—will be "bumped up" by thermal agitation to the level E_2. Because $E_2 > E_1$, the ratio $n(E_2)/n(E_1)$ will always be less than unity, which means that there will always be fewer atoms in the higher energy level than in the lower. This is what we would expect if the level populations are determined only by the action of thermal agitation.

If we expose a system like that of Fig. 48-20*a* to radiation, the dominant process—by sheer weight of numbers—will be absorption. However, if the level populations were inverted, as in Fig. 48-20*b*, the dominant process in the presence of radiation would be stimulated emission, and with it the generation of laser light. A *population inversion* like that of Fig. 48-20*b* is not a situation that is obtained by thermal processes; we must use clever tricks to bring it about.

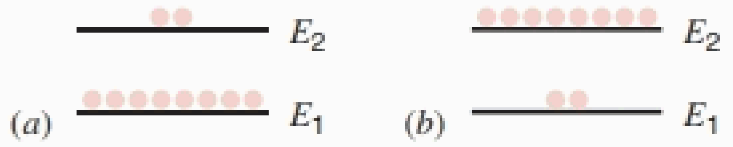


FIGURE 48-20. (*a*) The normal thermal equilibrium distribution of atomic systems occupying one or the other of two atomic states. (*b*) An inverted population, which can be obtained using special techniques.

How a Laser Works

Figure 48-21 shows schematically how a population inversion can be achieved so that laser action—or "lasing" as it is called—can occur. Atoms from the ground state E_1 are "pumped" up to an excited state E_3 —for example, by the absorption of energy from an intense light source that surrounds the lasing material. (In this section we discuss lasers whose operation depends on the properties of individual atoms in which electrons make transitions between well-defined states. In the next chapter we discuss semiconductor lasers, which operate on the basis of transitions between broad energy bands that are not associated with individual atoms.)

From E_3 the atoms decay rapidly to a state of energy E_2. For lasing to occur this state must be metastable; that is, it must have a relatively long mean life against decay by spontaneous emission. If conditions are right, state E_2 can then become more heavily populated than state E_1, thus providing the needed population inversion. A stray photon of the right energy can then trigger an avalanche of stimulated emission events, resulting in the production of laser light. A number of lasers using crystalline solids (such as ruby) as a lasing material operate in this three-level mode.

Figure 48-22 shows the elements of a common type of laser. The glass discharge tube is filled with an 80%–20% mixture of the inert gases helium and neon, with helium as the "pumping" medium and neon the "lasing" medium. Figure 48-23 is a simplified version of the level structures for these two atoms. Note that four levels, labeled E_0, E_1, E_2, and E_3, are involved in this lasing scheme, rather than three levels as in Fig. 48-21.

Pumping is accomplished by setting up an electrically induced gas discharge in the helium–neon mixture. Electrons and ions in this discharge occasionally collide with helium atoms, raising them to level E_3 in Fig. 48-23. This level is metastable, spontaneous emission to the ground

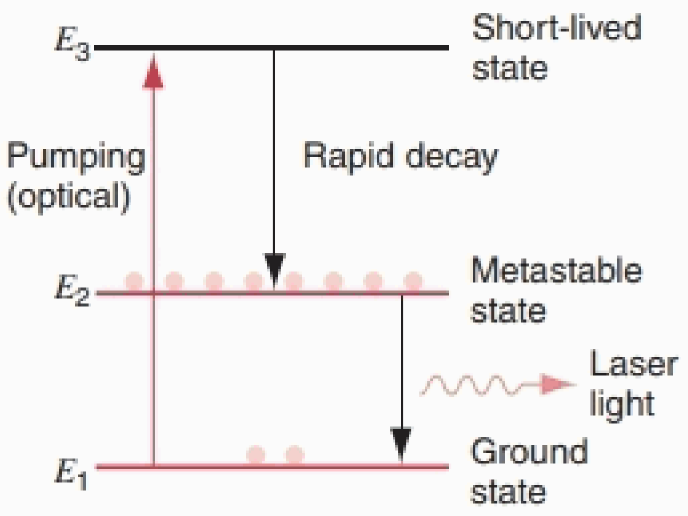


FIGURE 48-21. The basic three-level scheme for laser operation. Metastable state E_2 has a greater population than the ground state E_1; compare Fig. 48-20*b*.

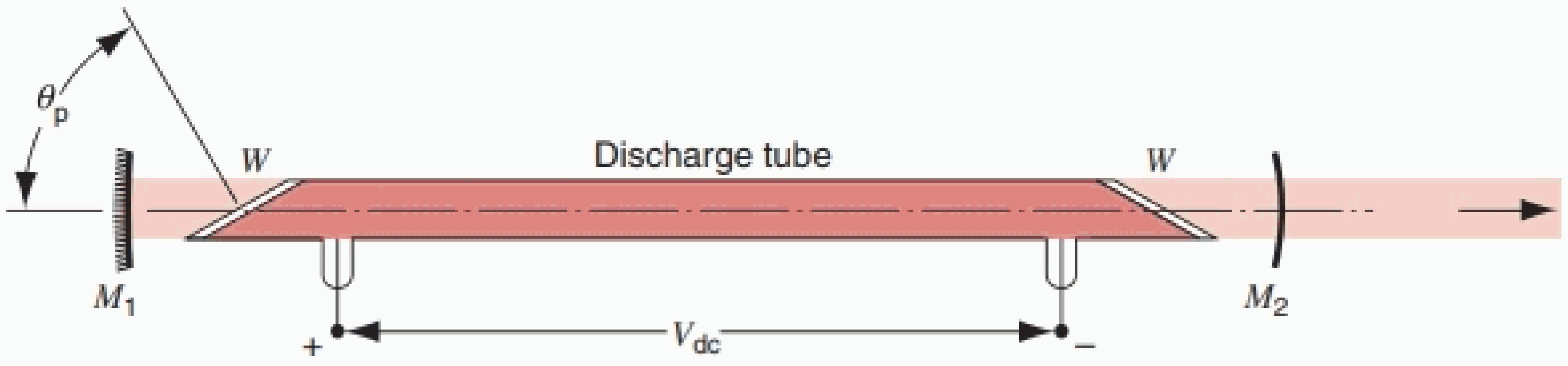


FIGURE 48-22. The basic elements of a helium–neon gas laser.

state (level E_0) being very infrequent. Level E_3 in helium ($= 20.61$ eV) is, by chance, very close to level E_2 in neon ($= 20.66$ eV), so that, during collisions between helium and neon atoms, the excitation energy of the helium can readily be transferred to the neon. In this way, level E_2 in Fig. 48-23 can become more highly populated than level E_1 in that figure. This population inversion is maintained because (1) the metastability of level E_3 ensures a ready supply of neon atoms in level E_2 and (2) level E_1 decays rapidly (through intermediate stages not shown) to the neon ground state, E_0. Stimulated emission from level E_2 to level E_1 predominates, and red laser light of wavelength 632.8 nm is generated.

Most stimulated emission photons initially produced in the discharge tube of Fig. 48-22 will not happen to be parallel to the tube axis and will be quickly stopped at the walls. Stimulated emission photons that *are* parallel to the axis, however, can move back and forth through the discharge tube many times by successive reflections from mirrors M_1 and M_2. These photons can in turn cause other stimulated emissions to occur. A chain reaction thus builds up rapidly in this direction, and the inherent parallelism of the laser light results.

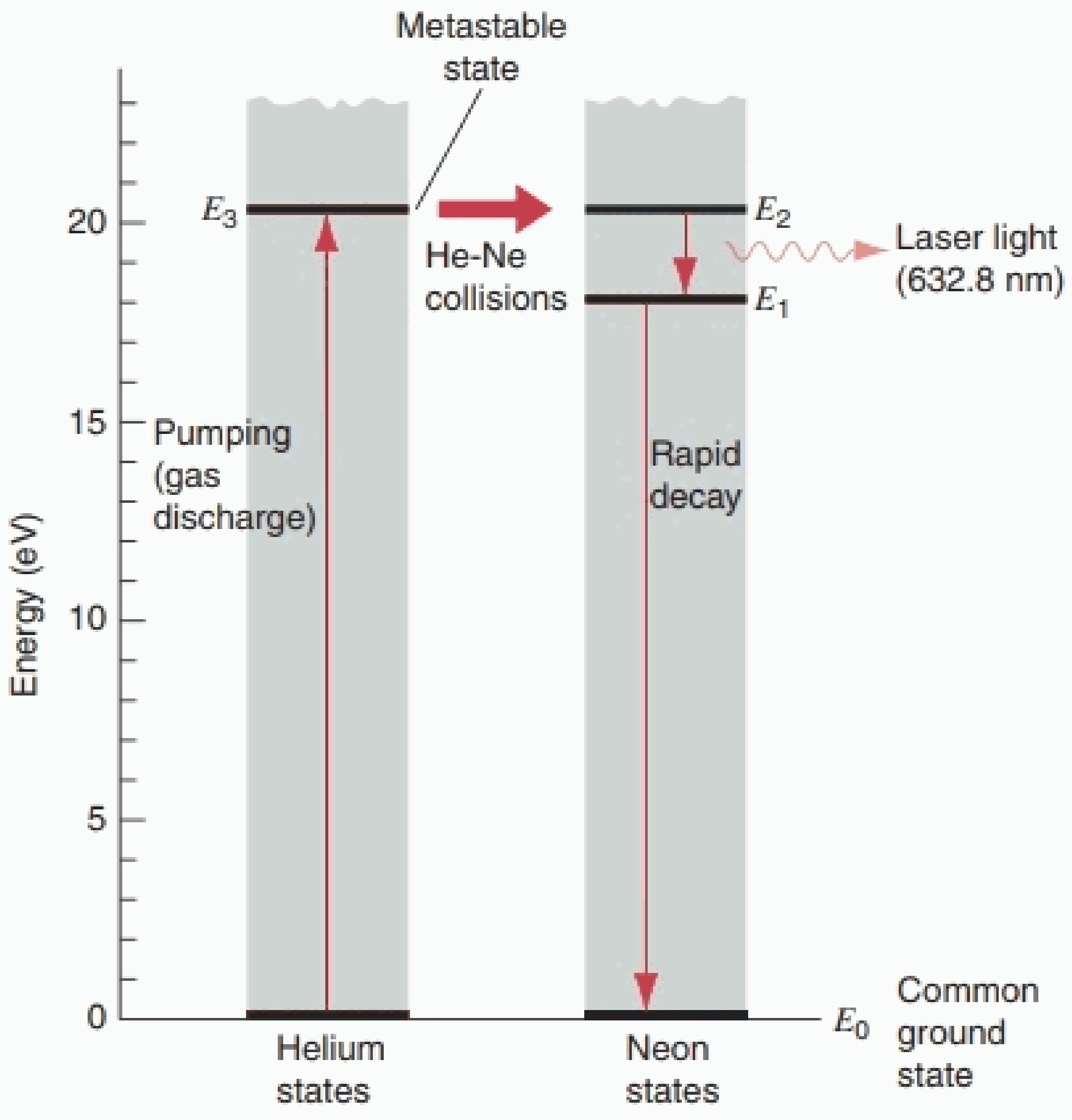


FIGURE 48-23. The atomic levels involved in the operation of a helium–neon gas laser.

Rather than thinking in terms of the photons bouncing back and forth between the mirrors, it is perhaps more useful to think of the entire arrangement of Fig. 48-22 as an optical resonant cavity that, like an organ pipe for sound waves, can be tuned to be sharply resonant at one (or more) wavelengths.

The mirrors M_1 and M_2 are concave, with their focal points nearly coinciding at the center of the tube. Mirror M_1 is coated with a dielectric film whose thickness is carefully adjusted to make the mirror as close as possible to totally reflective at the wavelength of the laser light; see Section 41-5. Mirror M_2, on the other hand, is coated so as to be slightly "leaky," so that a small fraction of the laser light can escape at each reflection to form the useful beam.

The windows *W* in Fig. 48-22, which close the ends of the discharge tube, are slanted so that their normals make an angle θ_p, the Brewster angle, with the tube axis, where

$$\tan \theta_p = n, \tag{48-20}$$

n being the index of refraction of the glass at the wavelength of the laser light. In Section 44-3 we showed that such windows transmit light without loss by reflection, provided only that the light is polarized with its plane of polarization in the plane of Fig. 48-22. If the windows were square to the tube ends, beam loss by reflection (about 4% from each surface of each window) would make laser operation impossible.

SAMPLE PROBLEM 48-7. A three-level laser of the type shown in Fig. 48-21 emits laser light at a wavelength of 550 nm, near the center of the visible band. (*a*) If the optical pumping mechanism is shut off, what will be the ratio of the population of the upper level (energy E_2) to that of the lower level (energy E_1)? Assume that $T = 300$ K. (*b*) At what temperature for the conditions of (*a*) would the ratio of populations be $\frac{1}{2}$?

Solution (*a*) From the Bohr frequency condition, the energy difference between the two levels is given by

$$E_2 - E_1 = hf = \frac{hc}{\lambda} = \frac{(6.63 \times 10^{-34}\ \text{J}\cdot\text{s})(3.00 \times 10^8\ \text{m/s})}{(550 \times 10^{-9}\ \text{m})(1.60 \times 10^{-19}\ \text{J/eV})}$$
$$= 2.26\ \text{eV}.$$

The mean energy of thermal agitation is equal to

$$kT = (8.62 \times 10^{-5}\ \text{eV/K})(300\ \text{K}) = 0.0259\ \text{eV}.$$

From Eq. 48-19, we have then, for the desired ratio,

$$n(E_2)/n(E_1) = e^{-(E_2 - E_1)/kT}$$
$$= e^{-(2.26\text{ eV})/(0.0259\text{ eV})} = e^{-87.3} = 1.3 \times 10^{-38}.$$

This is an incredibly small number. It is not unreasonable, however. An atom whose mean thermal agitation energy is only 0.0259 eV will not often impart an energy of 2.26 eV (87 times as great) to another atom in a collision.

(*b*) Setting the ratio in Eq. 48-19 equal to $\frac{1}{2}$, taking the natural logarithm of each side, and solving for T yields

$$T = \frac{E_2 - E_1}{k(\ln 2)} = \frac{2.26\text{ eV}}{(8.62 \times 10^{-5}\text{ eV/K})(0.693)} = 37{,}800\text{ K}.$$

This is much hotter than the surface of the Sun. It is clear that, if we are to invert the populations of these two levels, a special mechanism is needed. Without population inversion, lasing is not possible.

SAMPLE PROBLEM 48-8. A pulsed ruby laser has as its active element a synthetic ruby crystal in the form of a cylinder 6 cm long and 1 cm in diameter. Ruby consists of Al_2O_3 in which—in this case—one aluminum ion in every 3500 has been replaced by a chromium ion, Cr^{3+}. It is in fact the optical absorption properties of this small chromium "impurity" that account for the characteristic color of ruby. These same ions also account for the lasing ability of ruby, which occurs—by the three-level mechanism of Fig. 48-21—at a wavelength of 694.4 nm.

Suppose that *all* the Cr^{3+} ions are in a metastable state corresponding to state E_2 of Fig. 48-21 and that *none* are in the ground state E_1. How much energy is available for release in a single pulse of laser light if *all* these ions revert to the ground state in a single stimulated emission chain reaction episode? Our answer will be an upper limit only because the conditions postulated cannot be realized in practice. The density ρ of Al_2O_3 is 3700 kg/m³, and its molar mass M is 0.102 kg/mol.

Solution The number of Al^{3+} ions is

$$N_{\text{Al}} = \frac{2N_{\text{A}}m}{M} = \frac{2N_{\text{A}}\rho V}{M},$$

where m is the mass of the ruby cylinder and the factor 2 accounts for there being two aluminum ions in each "molecule" of Al_2O_3. The volume V is

$$V = (\pi/4)(1.0 \times 10^{-2}\text{ m})^2(6.0 \times 10^{-2}\text{ m}) = 4.7 \times 10^{-6}\text{ m}^3.$$

Thus

$$N_{\text{Al}} = \frac{(2)(6.0 \times 10^{23}/\text{mol})(3.7 \times 10^3\text{ kg/m}^3)(4.7 \times 10^{-6}\text{ m}^3)}{0.102\text{ kg/mol}}$$
$$= 2.1 \times 10^{23}.$$

The number of Cr^{3+} ions is then

$$N_{\text{Cr}} = \frac{N_{\text{Al}}}{3500} = 6.0 \times 10^{19}.$$

The energy of the stimulated emission photon is

$$E = hf = \frac{hc}{\lambda} = \frac{(4.1 \times 10^{-15}\text{ eV}\cdot\text{s})(3.0 \times 10^8\text{ m/s})}{694 \times 10^{-9}\text{ m}} = 1.8\text{ eV},$$

and the total available energy per laser pulse is

$$U = N_{\text{Cr}}E = (6.0 \times 10^{19})(1.8\text{ eV})(1.6 \times 10^{-19}\text{ J/eV}) = 17\text{ J}.$$

Such large pulse energies have indeed been achieved, but only by much more elaborate laser arrangements than that described here.

In this example we have postulated an ideal circumstance—namely, a total population inversion—in which the ground state remains virtually unpopulated. The actual population inversion in a working ruby laser will be very much less than total. For this and other reasons the pulse energy in practice will be very much less than the upper limit calculated above.

MULTIPLE CHOICE

48-1 The X-Ray Spectrum of Atoms

1. Which of the following x-ray lines will have the greatest frequency in a given element?
(A) K_α (B) K_β (C) L_α
(D) It depends on the element.

48-2 X Rays and the Numbering of the Elements

2. The K_α wavelength of an element with atomic number $Z = 17$ is given by λ.
(*a*) What is the atomic number of an element with a K_α wavelength of $\lambda/4$?
(A) 65 (B) 64 (C) 63 (D) 33 (E) 31
(*b*) What is the atomic number of an element with a K_α wavelength of 4λ?
(A) 3 (B) 4 (C) 5 (D) 7 (E) 9

3. The K_α wavelength of an element with atomic number Z is λ_Z. The K_α wavelength of an element with atomic number $2Z$ is λ_{2Z}. How are λ_Z and λ_{2Z} related?
(A) $\lambda_Z > 4\lambda_{2Z}$ (B) $\lambda_Z = 4\lambda_{2Z}$ (C) $\lambda_Z < 4\lambda_{2Z}$
(D) Depending on Z, λ_Z could be greater than or less than $4\lambda_{2Z}$.

4. According to the Bohr theory, how many electrons "screen" the charge of the nucleus as seen by an electron in the M shell?
(A) 2 (B) 6 (C) 7 (D) 9 (E) 10

48-3 Building Atoms

5. Answer this question *without* consulting a periodic table. An atom in the ground state has 17 electrons.
(*a*) What is the minimum number of shells that are occupied by at least some electrons for this atom?
(A) 1 (B) 2 (C) 3 (D) 4
(*b*) Assuming that *all* of the inner shells are completely filled, what is the number of electrons in the outermost shell?
(A) 2 (B) 5 (C) 7 (D) 8

6. Which of the following atoms would not have spherical symmetry in the ground state?
(A) He (B) Be (C) C (D) Ne

48-4 The Periodic Table

7. Answer this question by consulting a periodic table.
(*a*) Which of the following shows the correct ranking of the energies of the specified subshells?
(A) $E_{4p} < E_{4f} < E_{6s}$ (B) $E_{4f} < E_{4p} < E_{6s}$
(C) $E_{4p} < E_{6s} < E_{4f}$ (D) $E_{4f} < E_{6s} < E_{4p}$
(*b*) Which of the following shows the correct ranking of the energies of the specified subshells?
(A) $E_{5f} < E_{6d} < E_{7f}$ (B) $E_{5f} < E_{7s} < E_{6d}$
(C) $E_{7s} < E_{5f} < E_{6p}$ (D) $E_{7s} < E_{6d} < E_{5f}$

8. Which of the following transitions for an excited electron satisfies the selection rule (Eq. 48-6) and could result in the emission of a photon?
(A) $5p \rightarrow 6p$ (B) $7s \rightarrow 4p$
(C) $6s \rightarrow 5d$ (D) $3p \rightarrow 4f$

9. Consider the fictitious heavy element pandemonium, which has atomic number 169.
(*a*) In which period of the periodic table would this element probably appear?
(A) 8 (B) 9 (C) 10 (D) 11
(*b*) How many electrons would probably be in the outer shell of this element?
(A) 1 (B) 3 (C) 7 (D) 8

48-5 Atomic Magnetism

10. An iron sphere rests in a frictionless cradle inside a solenoid. The axis of the solenoid is in the north/south direction. The current in the solenoid is reversed; about which axis will the iron sphere begin to rotate?
(A) North/south (B) East/west (C) Up/down
(D) The sphere will not rotate, because it has spherical symmetry.

48-6 The Stern–Gerlach Experiment

11. Consider a hypothetical atom in which the valence electrons have a net spin of 0 and a net orbital angular momentum quantum number of 1 (p state). If a beam of these atoms passes through a Stern–Gerlach apparatus, into how many components would the beam split?
(A) 1 (B) 2 (C) 3 (D) 6

48-7 Nuclear Magnetic Resonance

48-8 Magnetism and Atomic Radiations

12. Consider electrons in the $4d$ subshell, which is split into two states according to Eq. 48-17.
(*a*) How many substates are present in the state with the larger value of j?
(A) 3 (B) 4 (C) 5 (D) 6
(*b*) How many substates are present in the state with the smaller value of j?
(A) 3 (B) 4 (C) 5 (D) 6

48-9 Lasers and Laser Light

QUESTIONS

1. ”If the angular momentum of electrons in atoms were not quantized, the periodic table of the elements would not be what it is.” Discuss this statement.

2. How would the properties of helium differ if the electron had no spin—that is, if the only operative quantum numbers were n, l, and m_l?

3. Explain how the interaction between the spin and the orbital motions of the valence electron in sodium leads to the splitting of the spectral lines of sodium, producing the familiar sodium doublet. See Fig. 48-15.

4. What is the origin of the cutoff wavelength $\lambda_{\min}$ of Fig. 48-1? Why is it an important clue to the photon nature of x rays?

5. In Fig. 48-2, why is the emitted photon shown moving off in the direction that it is? Could it be shown moving off in any other direction? Explain.

6. What are the characteristic x rays of an element? How can they be used to determine the atomic number of an element?

7. Compare Figs. 48-1 and 48-3. How can you be sure that the two prominent peaks in Fig. 48-1 do indeed correspond numerically with the two transitions similarly labeled in Fig. 48-3?

8. Can atomic hydrogen be caused to emit x rays? If so, describe how. If not, why not?

9. How does the x-ray energy-level diagram of Fig. 48-3 differ from the energy-level diagram for hydrogen that is displayed in Fig. 47-11? In what respects are the two diagrams similar?

10. When extended to higher atomic numbers, the Moseley plot of Fig. 48-4 is not a straight line but is concave upward. Does this affect the ability to assign atomic numbers to the elements?

11. Why is it that Bohr theory, which does not work very well even for helium ($Z = 2$), gives such a good account of the characteristic x-ray spectra of the elements, or at least of that portion that originates deep within the atom?

12. Why does the characteristic x-ray spectrum vary in a systematic way from element to element but the spectrum in the visible range does not?

13. Why do you expect the wavelengths of radiations generated by transitions deep within the atom to be shorter than those generated by transitions occurring in the outer fringes of the atom?

14. Given the characteristic x-ray spectrum of a certain element containing a number of lines, how would you go about identifying and labeling them?

15. On what quantum numbers does the energy of an electron in (*a*) a hydrogen atom and (*b*) a vanadium atom depend?

16. The periodic table of the elements was based originally on atomic mass rather than on atomic number, the latter concept having not yet been developed. Why were such early tables as successful as they proved to be? In other words, why is the atomic mass of an element (roughly) proportional to its atomic number?

17. How does the structure of the periodic table support the need for a fourth quantum number, corresponding to electron spin?

18. If there were only three quantum numbers (that is, if the electron had no spin), how would the chemical properties of helium be different?

19. Explain why the effective radius of a helium atom is less than that of a hydrogen atom.

20. Why does it take more energy to remove an electron from neon ($Z = 10$) than from sodium ($Z = 11$)?

21. What can Fig. 48-5 tell you about why the inert gases are so chemically stable?

22. Does it make any sense to assign quantum numbers to a vacancy in an otherwise filled subshell?

23. Why do the lanthanide series of elements (see Appendix E) have such similar chemical properties? How can we justify putting them all into a single square of the periodic table? Why is it that, in spite of their similar chemical properties, they can be so easily sorted out by measuring their characteristic x-ray spectra?

24. In your own words, state the minimum energy principle for atom building and give a physical argument in support of it.

25. Figure 48-5 shows that the 2*s* state is lower in energy than the 2*p* state. Can you explain why this should be so, basing your argument on the radial probability densities of the two states (see Figs. 47-14 and 47-15)?

26. If you start with a bare nucleus and fill in the electrons to form an atom in its ground state, the energies of the unfilled levels change as you proceed. Why do they change? Do they increase or decrease in energy as electrons are added?

27. Justify the statement that, in the Einstein–de Haas effect, the angular momentum of the iron bar as a whole must be conserved when the bar is suddenly magnetized.

28. In the Einstein–de Haas experiment (see Fig. 48-8), can you justify the fact that the predicted period of rotation of the cylinder depends only on the cylinder radius and not, for example, on its height? What assumptions were made in deriving the expression for the period of rotation?

29. Convince yourself that the directions of the arrows in Fig. 48-8*b* representing the current in the solenoid, the magnetic field, the atomic angular momenta, and the direction of rotation of the cylinder are consistent with each other.

30. Does the Einstein–de Haas effect provide any evidence that angular momentum is quantized?

31. A beam of circularly polarized light, viewed as a beam of photons whose spins are aligned, can exert a torque on an absorbing screen. Develop the analogy to the Einstein–de Haas experiment.

32. A beam of neutral silver atoms is used in a Stern–Gerlach experiment. What is the origin of both the force and the torque that act on the atom? How is the atom affected by each?

33. What determines the number of subbeams into which a beam of neutral atoms is split in a Stern–Gerlach experiment?

34. If in a Stern–Gerlach experiment an ion beam is resolved into five component beams, then what angular momentum quantum number does each ion have?

35. In a Stern–Gerlach apparatus, is it possible to have a magnetic field configuration in which the magnetic field itself is zero along the beam path but the field gradient is not? If your answer is yes, can you design an electromagnet that will produce such a field configuration?

36. The silver atoms in the Stern–Gerlach experiment of Sample Problem 48-4 are uncharged. Suppose that a silver atom in the apparatus were suddenly to lose an electron, becoming a silver ion. What would be the nature and the relative magnitude of the forces acting on it (*a*) before and (*b*) after this event?

37. How would you show in the laboratory that an atom has angular momentum? That it has a magnetic dipole moment?

38. In the hydrogen atom state with $l = 1$, the spin and the orbital angular momentum vectors can be aligned either parallel or antiparallel. Which arrangement has the greater energy and why?

39. How do we arrive at the conclusion that the spin magnetic quantum number m_s can have only the values $\pm\frac{1}{2}$? What kinds of experiments support this conclusion?

40. Why is the magnetic moment of the spinning electron directed opposite to its spin angular momentum?

41. Why is focused laser light inherently better than focused light from a tiny incandescent lamp filament for delicate surgical jobs such as spot-welding detached retinas?

42. Laser light forms an almost parallel beam. Does the intensity of such light fall off as the inverse square of the distance from the source?

43. In what ways are laser light and star light similar? In what ways are they different?

44. Arthur Schawlow, one of the laser pioneers, invented a typewriter eraser based on focusing laser light on the unwanted character. Can you imagine what its principle of operation is?

45. In what ways do spontaneous emission and stimulated emission differ?

46. We have spontaneous emission and stimulated emission. From symmetry, why do we not also have spontaneous and stimulated absorption? Discuss in terms of Fig. 48-19.

47. Why is a population inversion necessary between two atomic levels for laser action to occur?

48. What is a metastable state? What role do such states play in the operation of a laser?

49. Comment on this statement: "Other things being equal, a four-level laser scheme such as that of Fig. 48-23 is preferable to a three-level scheme such as that of Fig. 48-21 because, in the three-level scheme, one-half of the population of atoms in level E_1 must be moved to state E_2 before a population inversion can even begin to occur."

50. Comment on this statement: "In the laser of Fig. 48-22, only light whose plane of polarization lies in the plane of that figure is transmitted through the right-hand window. Therefore, half of the energy potentially available is lost." (Hint: Is this second statement really true? Consider what happens to photons whose effective plane of polarization is at right angles to the plane of Fig. 48-22. Do such photons participate fully in the stimulated emission amplification process?)

51. A beam of light emerges from an aperture in a "black box" and moves across your laboratory bench. How could you test this beam to find out the extent to which it is coherent over its cross section? How could you tell (without opening the box) whether or not the concealed light source is a laser?

52. Why is it difficult to build an x-ray laser?

EXERCISES

48-1 The X-Ray Spectrum of Atoms

1. Show that the short-wavelength cutoff in the continuous x-ray spectrum is given by

$$\lambda_{\min} = 1240 \text{ pm}/\Delta V,$$

where ΔV is the applied potential difference in kilovolts.

2. Determine Planck's constant from the fact that the minimum x-ray wavelength produced by 40.0-keV electrons is 31.1 pm.

3. What is the minimum potential difference across an x-ray tube that will produce x rays with a wavelength of 0.126 nm?

4. In Fig. 48-1, the x rays shown are produced when 35.0-keV electrons fall on a molybdenum target. If the accelerating potential is maintained at 35.0 kV but a silver target ($Z = 47$) is substituted for the molybdenum target, what values of (*a*) $\lambda_{\min}$, (*b*) λ_{K_β} and (*c*) λ_{K_α} result? The *K*, *L*, and *M* atomic levels for silver (compare with Fig. 48-3) are 25.51, 3.56, and 0.53 keV, respectively.

5. Electrons bombard a molybdenum target, producing both continuous and characteristic x rays as in Fig. 48-1. In that figure the energy of the incident electrons is 35.0 keV. If the accelerating potential applied to the x-ray tube is increased to 50.0 kV, what values of (*a*) $\lambda_{\min}$, (*b*) λ_{K_β}, and (*c*) λ_{K_α} result?

6. The wavelength of the K_α line from iron is 19.3 pm. (*a*) Find the energy difference between the two states of the iron atom (see Fig. 48-3) that give rise to this transition. (*b*) Find the corresponding energy difference for the hydrogen atom. Why is the difference so much greater for iron than for hydrogen? (Hint: In the hydrogen atom the *K* shell corresponds to $n = 1$ and the *L* shell to $n = 2$.)

7. From Fig. 48-1, calculate approximately the energy difference $E_L - E_M$ for the x-ray atomic energy levels of molybdenum. Compare with the result that may be found from Fig. 48-3.

8. Find the minimum potential difference that must be applied to an x-ray tube to produce x rays with a wavelength equal to the Compton wavelength of the electron. (See Exercise 30 of Chapter 45.)

9. X rays are produced in an x-ray tube by a target potential of 50.0 kV. If an electron makes three collisions in the target before coming to rest and loses one-half of its remaining kinetic energy on each of the first two collisions, determine the wavelengths of the resulting photons. Neglect the recoil of the heavy target atoms.

10. A tungsten target ($Z = 74$) is bombarded by electrons in an x-ray tube. (*a*) What is the minimum value of the accelerating potential that will permit the production of the characteristic K_β and K_α lines of tungsten? (*b*) For this same accelerating potential, what is the value of $\lambda_{\min}$? (*c*) Calculate λ_{K_β} and λ_{K_α}. The *K*, *L*, and *M* atomic levels for tungsten (see Fig. 48-3) are 69.5, 11.3, and 2.3 keV, respectively.

11. A molybdenum target ($Z = 42$) is bombarded with 35.0-keV electrons and the x-ray spectrum of Fig. 48-1 results. Here $\lambda_{K_\beta} = 63$ pm and $\lambda_{K_\alpha} = 71$ pm. (*a*) What are the corresponding photon energies? (*b*) It is desired to filter these radiations through a material that will absorb the K_β line much more strongly than it will absorb the K_α line. What substance(s) would you use? The *K* ionization energies for molybdenum and for four neighboring elements are as follows:

Z	40	41	42	43	44
Element	Zr	Nb	Mo	Tc	Ru
E_K (keV)	18.00	18.99	20.00	21.04	22.12

(Hint: A substance will selectively absorb one of two x radiations more strongly if the photons of one have enough energy to eject a *K* electron from the atoms of the substance but the photons of the other do not.)

12. The binding energies of *K*-shell and *L*-shell electrons in copper are 8.979 keV and 0.951 keV, respectively. If a K_α x ray from copper is incident on a sodium chloride crystal and gives a first-order Bragg reflection at 15.9° when reflected from the alternating planes of the sodium atoms, what is the spacing between these planes?

48-2 X Rays and the Numbering of the Elements

13. Here are the K_α wavelengths of a few elements:

Ti	27.5 pm	Co	17.9 pm
V	25.0 pm	Ni	16.6 pm
Cr	22.9 pm	Cu	15.4 pm
Mn	21.0 pm	Zn	14.3 pm
Fe	19.3 pm	Ga	13.4 pm

Make a Moseley plot (see Fig. 48-4) and verify that its slope agrees with the value calculated in Sample Problem 48-2.

14. Using the Bohr theory, calculate the ratio of the wavelengths of the K_α line for niobium (Nb) to that of gallium (Ga). Take needed data from the periodic table.

48-3 Building Atoms

48-4 The Periodic Table

15. Two electrons in lithium ($Z = 3$) have as their quantum numbers n, l, m_l, m_s, the values 1, 0, 0, $\pm\frac{1}{2}$, respectively. (*a*) What quantum numbers can the third electron have if the atom is to be in its ground state? (*b*) If the atom is to be in its first excited state?

16. If a uranium nucleus ($Z = 92$) had only a single electron, what would be the radius of its ground-state orbit, according to Bohr's theory?

17. If the electron had no spin, and if the Pauli exclusion principle still held, how would the periodic table be affected? In particular, which of the present elements would be inert gases?

18. By inspection of Fig. 48-5, what do you think might be the atomic number of the next higher inert gas above $Z = 118$? Assume that the 5*g* subshell appears in the 8th period.

19. In the alkali metals there is one electron outside a closed shell. (*a*) Using the Bohr theory, calculate the effective charge number of the nucleus as seen by the valence electron in sodium (ionization energy = 5.14 eV) and potassium (ionization energy = 4.34 eV). (*b*) For each element, what fraction is this of the actual nuclear charge Z? Needed quantum numbers can be found on Fig. 48-5.

20. Suppose there are two electrons in the same system, both of which have $n = 2$ and $l = 1$. (*a*) If the exclusion principle did not apply, how many combinations of states are conceivably possible? (*b*) How many states does the exclusion principle forbid? Which ones are they?

48-5 Atomic Magnetism

48-6 The Stern–Gerlach Experiment

21. (*a*) Show that the magnetic moments of the electrons in the various Bohr orbits are given, according to the Bohr theory, by

$$\mu = n\mu_B,$$

in which μ_B is the Bohr magneton and $n = 1, 2, 3, \ldots$. (*b*) How does this expression compare with the actual values?

22. Suppose a hydrogen atom (in its ground state) moves 82 cm in a direction perpendicular to a magnetic field that has a gradient, in the vertical direction, of 16 mT/m. (*a*) What is the force on the atom due to the magnetic moment of the electron, which we take to be 1 Bohr magneton? (*b*) Find its vertical displacement if its speed is 970 m/s.

23. Calculate the acceleration of the silver atom as it passes through the deflecting magnet in the Stern–Gerlach experiment of Sample Problem 48-5.

24. Assume that in the Stern–Gerlach experiment described for neutral silver atoms the magnetic field $\vec{\mathbf{B}}$ has a magnitude of 520 mT. (*a*) What is the energy difference between the orientations of the silver atoms in the two subbeams? (*b*) What is the frequency of the radiation that would induce a transition between these two states? (*c*) What is its wavelength, and to what part of the electromagnetic spectrum does it belong? The magnetic moment of a neutral silver atom is 1 Bohr magneton.

25. What is the wavelength of a photon that will induce a transition of an electron spin from parallel to antiparallel orientation in a magnetic field of magnitude 190 mT?

48-7 Nuclear Magnetic Resonance

48-8 Magnetism and Atomic Radiations

26. The proton as well as the electron has spin 1/2. In the hydrogen atom in its ground state there are two energy levels, depending on whether the electron and the proton spins are in the same direction or in opposite directions. The state with the spins in the opposite direction has the higher energy. If an atom is in this state and one of the spins "flips over," the small energy difference is released as a photon of wavelength 21 cm. This spontaneous spin-flip process is very slow, the mean life for the process being about 10^7 y. However, radio astronomers observe this 21-cm radiation from interstellar space, where the density of hydrogen is so small that an atom can flip before being disturbed by collisions with other atoms. Calculate the effective magnetic field (due to the magnetic dipole moment of the proton) experienced by the electron in the emission of this 21-cm radiation.

48-9 Lasers and Laser Light

27. Lasers have become very small as well as very large. The active volume of a laser constructed of the semiconductor GaAlAs has a volume of only 200 $(\mu\text{m})^3$ (smaller than a grain of sand) and yet it can continuously deliver 5.0 mW of power at 0.80-μm wavelength. Calculate the production rate of photons.

28. A ruby laser emits light at wavelength 694.4 nm. If a laser pulse is emitted for 12.0 ps and the energy release per pulse is 150 mJ, (*a*) what is the length of the pulse, and (*b*) how many photons are in each pulse?

29. It is entirely possible that techniques for modulating the frequency or amplitude of a laser beam will be developed so that such a beam can serve as a carrier for television signals, much as microwave beams do now. Assume also that laser systems will be available whose wavelengths can be precisely "tuned" to anywhere in the visible range—that is, in the range 400 nm $< \lambda <$ 700 nm. If a television channel occupies a bandwidth of 10 Mhz, how many channels could be accommodated with this laser technology? Comment on the intrinsic superiority of visible light to microwaves as carriers of information.

30. A He-Ne laser emits light at a wavelength of 632.8 nm and has an output power of 2.3 mW. How many photons are emitted each minute by this laser when operating?

31. A hypothetical atom has energy levels evenly spaced by 1.2 eV in energy. For a temperature of 2000 K, calculate the ratio of the number of atoms in the 13th excited state to the number in the 11th excited state.

32. A population inversion for two levels is often described by assigning a negative Kelvin temperature to the system. Show that such a negative temperature would indeed correspond to an inversion. What negative temperature would describe the system of Sample Problem 48-7 if the population of the upper level exceeds that of the lower by 10.0%?

33. An atom has two energy levels with a transition wavelength of 582 nm. At 300 K, 4.0×10^{20} atoms are in the lower state. (*a*) How many occupy the upper state under conditions of thermal equilibrium? (*b*) Suppose, instead, that 7.0×10^{20} atoms are pumped into the upper state, with 4.0×10^{20} in the lower state. How much energy could be released in a single laser pulse?

34. The mirrors in the laser of Fig. 48-22 form a cavity in which standing waves of laser light are set up. In the vicinity of 533 nm, how far apart in wavelength are the adjacent allowed operating modes? The mirrors are 8.3 cm apart.

35. The beam from an argon laser (λ = 515 nm) has a diameter d of 3.00 mm and a power output of 5.21 W. The beam is focused onto a diffuse surface by a lens of focal length f = 3.50 cm. A diffraction pattern such as that of Fig. 42-13 is formed. (*a*) Show that the radius of the central disk is given by

$$R = \frac{1.22 f\lambda}{d}.$$

The central disk can be shown to contain 84% of the incident power. Calculate (*b*) the radius R of the central disk, and the average power flux density (*c*) in the incident beam and (*d*) in the central disk.

36. The use of lasers for defense against ballistic missiles is being studied. A laser beam of intensity 120 MW/m^2 would probably burn into and destroy a hardened (nonspinning) missile in about 1 s. (*a*) If the laser has a power output of 5.30 MW, a wavelength of 2.95 pm, and a beam diameter of 3.72 m (a very powerful laser indeed), would it destroy a missile at a distance of 3000 km? (*b*) If the wavelength could be changed, what minimum value would work? (*c*) If the wavelength of the laser could not be changed, what would be the destructive range of the laser in (*a*)? Use the equation for the central disk given in Exercise 35 and take the focal length to be the distance to the target.

PROBLEMS

1. A 20.0-keV electron is brought to rest by undergoing two successive bremsstrahlung events, thus transferring its kinetic energy into the energy of two photons. The wavelength of the second photon is 130 pm greater than the wavelength of the first photon to be emitted. (*a*) Find the energy of the electron after its first deceleration. (*b*) Calculate the wavelengths and energies of the two photons.

2. In an x-ray tube, an electron moving initially at a speed of 2.73×10^8 m/s slows down in passing near a nucleus. A single photon of energy 43.8 keV is emitted. Find the final speed of the electron. (Relativity must be taken into account; ignore the energy imparted to the nucleus.)

3. Show that a moving electron cannot spontaneously emit an x-ray photon in free space. A third body (atom or nucleus) must be present. Why is it needed? (Hint: Examine conservation of total energy and of momentum.)

4. Let λ_α and λ_β represent respectively the wavelengths of the K_α and K_β x rays of an element. (*a*) Show that the Bohr theory gives

$$\frac{\lambda_\beta}{\lambda_\alpha} = k\,\frac{(Z-1)^2}{Z^2-9}$$

and find the numerical value of k. (*b*) Compare the predicted value of this ratio with that determined from the x-ray wavelengths of some elements with small, medium, and large values of Z. (X-ray wavelengths are tabulated in the *Handbook of Chemistry and Physics*.)

5. The two wavelengths that appear in the $3p \rightarrow 3s$ transition in sodium (the so-called sodium *D* lines; see Figs. 48-7 and 48-15) are due to the fine-structure splitting of the $3p$ level. (*a*) From the wavelengths of the two members of the $3p \rightarrow 3s$ doublet given in Fig. 48-15, find the energy splitting of the two fine-structure components of the $3p$ level. (*b*) Use the energy splitting found in (*a*) to predict the difference in wavelength of the two members of the $4s \rightarrow 3p$ doublet shown in Fig. 48-7.

6. (*a*) Using the Bohr theory, derive an expression for the radius of the *K* shell of an alkali element in terms of its atomic number Z. (*b*) Again using the Bohr theory, derive an expression for the radius of the outermost shell of an alkali element, and show that it is independent of Z.

7. A crude generalization of the Bohr theory for helium is to assume that the two electrons are in the same orbit, but always on opposite sides of the nucleus. (*a*) Show that the ground state orbit would have a radius of $4a_0/7$, where a_0 is the Bohr radius. (*b*) Find the energy of the atom in this state. (*c*) What would be the ionization energy of this atom (the energy needed to remove one electron)? Compare with the measured ionization energy of helium (24.6 eV).

8. A particular (hypothetical) atom has only two atomic levels, separated in energy by 3.2 eV. In the atmosphere of a star there are 6.1×10^{13} of these atoms per cm^3 in the excited (upper) state and 2.5×10^{15} per cm^3 in the ground (lower) state. Calculate the temperature of the star's atmosphere.

9. A high-powered laser beam ($\lambda = 600$ nm) with a beam diameter of 11.8 cm is aimed at the Moon, 3.82×10^5 km distant. The spreading of the beam is caused only by diffraction effects. The angular location of the edge of the central diffraction disk (see Eq. 42-11) is given by

$$\sin\theta = \frac{1.22\lambda}{d},$$

where d is the diameter of the beam aperture. Find the diameter of the central diffraction disk at the Moon's surface.

10. The active medium in a particular ruby laser ($\lambda = 694$ nm) is a synthetic ruby crystal 6.00 cm long and 1.0 cm in diameter. The crystal is silvered at one end and—to permit the formation of an external beam–only partially silvered at the other. (*a*) Treat the crystal as an optical resonant cavity in analogy to a closed organ pipe and calculate the number of standing-wave nodes there are along the crystal axis. (*b*) By what amount Δf would the beam frequency have to shift to increase this number by one? Show that Δf is just the inverse of the travel time of light for one round trip back and forth along the crystal axis. (*c*) What is the corresponding fractional frequency shift $\Delta f/f$? The appropriate index of refraction is 1.75.

CHAPTER 49

ELECTRICAL CONDUCTION IN SOLIDS

We have seen in the previous two chapters how well quantum theory works when we apply it to individual atoms. In this chapter we show that this powerful theory works equally well when we apply it to collections of atoms in the form of solids.

Every solid has an enormous range of properties that we can choose to examine: Is it soft or hard? Can it be hammered into a thin sheet or drawn into a fine wire? Is it transparent? What kind of waves travel through it and at what speeds? Does it conduct heat? What are its magnetic properties? What is its crystal structure? And so on. In each case, we should like to use quantum theory to understand the measured properties.

In this chapter we focus on one particular property of solids: conduction of electricity. We discuss the classification of solids into conductors, insulators, semiconductors, and superconductors, and we show how quantum theory provides the framework for understanding why some materials behave one way and some another.

49-1 QUANTUM THEORY OF SOLIDS

In the previous two chapters, we have seen how quantum mechanics has been successfully applied to the analysis of the structure of atoms. Based on the laws of quantum mechanics and the Coulomb potential, we can understand many properties of atoms (size, radiations they emit and absorb, angular momentum, magnetic moments, and so on). The periodic table, which was first developed by arranging atoms according to their common properties and their mass, then becomes a direct outcome of the quantum theory of atoms.

We can develop a successful theory of atomic structure based on the properties of a single electron or at most a few valence electrons. This theory works very well for analyzing the properties of the atoms in a gas, because the electrons are fairly well localized to individual atoms, and each atom's properties depend only on the arrangement of its particular electrons.

In this chapter we will apply quantum theory to the structure of solids in order to understand their electrical conductivity. We will find that there are a number of differences between the application of quantum theory to individual atoms and its application to solids. First and foremost, the electrons that contribute most to the electrical conductivity in a solid do not "belong to" any particular atom; they are shared among all the atoms in the material. As a consequence, we may be dealing with very large numbers of electrons, so it makes no sense to try to follow the motion of an individual electron. Instead, we use statistical methods, as we did in Chapter 22 in analyzing the speed distribution of the molecules in a gas. The joining of quantum theory and statistical mechanics is called *quantum statistical mechanics.* We will use the methods of quantum statistical mechanics in analyzing the conductivity of solids.

Another difference between a solid and a gas is that in a solid the atoms interact with one another. In previous chapters we assumed that we were dealing with single, isolated atoms, in which the electrons interact only with the nucleus and the other electrons of their "home" atom. In a solid, the atoms are so close together that the electron energy levels in one atom are affected by the presence of the other atoms. This has the effect of changing the sharp energy levels of an isolated atom into broad energy bands. The bands are in reality collections of energy levels that are so numerous and so close together that we can regard them as continuous.

One similarity between the quantum theory of atoms and of solids is that in both cases the electrons must obey the Pauli exclusion principle, except that here we apply it to the entire solid instead of to only a single atom. In the entire solid, no two electrons can have the same set of quantum numbers. Electrons that are attached to different atoms in the solid must have different labels to identify their "home" atom, but what about the electrons that "belong to" the entire solid instead of to individual atoms? In this chapter we are mainly concerned with analyzing these electrons, because the electrons that are attached to individual atoms do not normally contribute to electrical conduction. Our goal then is to apply the Pauli principle to a collection of perhaps 10^{22} electrons, which would surely be a formidable task were it not for the methods of quantum statistics.

49-2 CONDUCTION ELECTRONS IN A METAL

An isolated copper atom has 29 electrons. In solid copper, 28 of these are held close to their lattice sites by electromagnetic forces and are not free to move throughout the volume of the solid. The remaining electron *is* free to so move and, if we apply an emf between the ends of a copper wire, it is these *conduction electrons* (one per atom) that constitute the current that is set up in the wire.

In Section 29-5 we looked at this problem from the point of view of classical physics, comparing the conduction electrons in a metal cube to the atoms of a gas confined to a cubical box. Using this (classical) *free electron gas* model, we derived an expression for the resistivity of the metal. It is (see Eq. 29-20)

$$\rho = \frac{m}{ne^2\tau}, \tag{49-1}$$

in which m is the mass and e the charge of the electron, n is the number of conduction electrons per unit volume, and τ is the average time between collisions of the electrons with the lattice.

We showed in Section 29-5 that τ is essentially constant, independent of whether or not an electric field has been set up inside the cube by an externally applied emf. Thus the resistivity ρ is independent of the applied electric field, which is another way of saying that metals obey Ohm's law.

Although this derivation of the form of Ohm's law is a fine achievement for classical physics, it is no simple matter to go much further. Also, there are some physical properties—the heat capacities of metals, for example—about which this classical theory *does* have something to say, but unfortunately its predictions do not agree with experiment. Looking beyond this level of concern, it is hard to imagine how we would explain something as complicated as a transistor on the basis of the classical free-electron gas model. We had better see what quantum mechanics has to offer.

The first step in solving any quantum mechanical problem is to specify the potential energy of the particle—which we take to be a single conduction electron—as a function of its position. We need this information to substitute into the Schrödinger equation. We start with the simplest reasonable assumption—namely, that the potential energy is zero for all points within the cubical metal sample and that it approaches an infinitely great value for all points outside. We are still dealing with a free electron gas, but it is now one that is governed by quantum—rather than classical—rules.

This potential energy reminds us of the problem of an electron trapped in an infinite well that we solved in Section 47-2. Note, however, two differences: the present problem is three-dimensional, and it involves a well of macroscopic, rather than atomic, dimensions.

We represent a single conduction electron trapped in its metal cube by a (standing) matter wave $\psi(\vec{\mathbf{r}})$, in which $\vec{\mathbf{r}}$ is a position vector, and we impose the condition that the probability density $\psi^2(\vec{\mathbf{r}})$ be zero both at the surface of the cube and at all outside points. This is our way of recognizing that the electron is truly trapped inside the metal cube. Figure 47-4 reminds us that we proceeded in the same way in the one-dimensional case.

If we impose these boundary conditions on the wave function, the Schrödinger equation tells us that the total energy E of the electron will be quantized, just as it was for an electron trapped in a one-dimensional well. There is a big difference, however. Because our metal cube is so very large on the scale of atomic dimensions, the number of standing matter waves that we can fit into the volume of the cube and still satisfy the boundary requirements is enormous, and the allowed electron energies are extremely close together. Sample Problem 49-1 shows that, for a cube 1 cm on edge, there are about 10^{20} quantized states that lie between $E = 5$ eV and $E = 5.01$ eV! Compare this with the limited array of well-spaced levels shown, for example, for the finite potential well in Fig. 47-6*b*.

We cannot possibly deal with this vast number of states one at a time; we must use statistical methods. Instead of asking, "What is the energy of this state?" we must ask, "How many states have energies that lie in the range E to $E + dE$?"

We have met situations like this before. For example, in describing the speeds of the molecules of an ideal gas in Section 22-4, we saw that the only way to proceed was to pose the question: "How many molecules have speeds that lie in the range v to $v + dv$?"

For the conduction electrons, the number of states (per unit volume of the solid) whose energies lie in the range E to $E + dE$ can be written as $n(E)dE$, where $n(E)$ is a function called the *density of states.* For our (quantum) free-electron gas it can be shown to be*

$$n(E) = \frac{8\sqrt{2}\pi m^{3/2}}{h^3} E^{1/2}. \qquad (49\text{-}2)$$

At this stage we are simply counting the states that are available to a single conduction electron. Note that there is nothing in Eq. 49-2 that depends on the material of which our sample is made. Fitting patterns of standing waves into a cubical box is a purely geometrical problem. In the following section we shall see how to go about filling those states.

You may well ask, "If the energies of the allowed states are so close together, why don't we just forget about the quantization and assume a continuous distribution in energy?" The answer, as we shall see in the next section, rests on the fact that the Pauli exclusion principle applies to electrons wherever we find them, whether as orbital electrons in atoms or as conduction electrons in metals. Even though there are many states in our problem, there are also many conduction electrons available to occupy them, and the Pauli principle allows us to put only one electron in each of these states. Thus, even though we cannot easily detect directly the quantized nature of the energies of the conduction electrons, the fact of quantization remains an absolutely central feature and has important consequences.

SAMPLE PROBLEM 49-1. A cube of copper is 1 cm on edge. How many states are available for its conduction electrons in the energy interval between $E = 5.00$ eV and 5.01 eV? Assume that the conduction electrons behave like a (quantum) free-electron gas.

Solution These energy limits are so close together that we can safely say that the answer, on a per unit volume basis, is $n(E)\Delta E$, where $E = 5$ eV and $\Delta E = 0.01$ eV. From Eq. 49-2 we have

$$\begin{aligned} n(E) &= \frac{8\sqrt{2}\pi m^{3/2}}{h^3} E^{1/2} \\ &= \frac{(8\sqrt{2}\pi)(9.11 \times 10^{-31}\ \mathrm{kg})^{3/2}}{(6.63 \times 10^{-34}\ \mathrm{J \cdot s})^3} [(5\ \mathrm{eV})(1.6 \times 10^{-19}\ \mathrm{J/eV})]^{1/2} \\ &= 9.48 \times 10^{46}\ \mathrm{m^{-3}\,J^{-1}} = 1.52 \times 10^{28}\ \mathrm{m^{-3}eV^{-1}}. \end{aligned}$$

Note that we must express the energy E in joules before substituting into Eq. 49-2, even though we wish our final result to be given in terms of electron volts.

The actual number N of states that lie in the range from $E =$ 5.00 eV to $E = 5.01$ eV in our cube is, if a is the length of the cube edge,

$$\begin{aligned} N &= n(E)\Delta E\, a^3 \\ &= (1.52 \times 10^{28}\ \mathrm{m^{-3}eV^{-1}})(0.01\ \mathrm{eV})(1 \times 10^{-2}\ \mathrm{m})^3 \\ &= 1.52 \times 10^{20}. \end{aligned}$$

That is, there are 1.52×10^{20} individual energy states between 5.00 eV and 5.01 eV. The average energy interval ΔE_{adj} between adjacent levels in this interval follows readily from

$$\Delta E_{\mathrm{adj}} = \frac{\Delta E}{N} = \frac{0.01\ \mathrm{eV}}{1.52 \times 10^{20}} \approx 7 \times 10^{-23}\ \mathrm{eV}.$$

We conclude that, even in this narrow energy band, there are very many states and they lie exceedingly close together in energy.

Our conclusions are completely independent of the material of the sample. Nor is it important that the sample is cubical; any other shape enclosing the same volume would give the same final result. What we *have* assumed to be true is that the conduction electrons behave like a (quantum) free-electron gas. That is, we have assumed that their potential energy is constant (which we have taken to be zero) for all points within the sample. For actual metals this assumption is never strictly true. Nevertheless, our central conclusion holds: many states are available to a conduction electron in a metal and they lie very close together in energy.

49-3 FILLING THE ALLOWED STATES

Now that we have seen how many states there are, we are ready to start filling them with electrons. We went through this process in Section 48-4 in connection with building up the periodic table of the elements. There we saw the central importance of Pauli's exclusion principle, which tells us that we can allocate only one electron to a given state. This powerful principle is just as important for our present problem.

Figure 49-1*a* shows the density of states given by Eq. 49-2. This function gives the number of possible states in any energy interval. However, not all those states are occupied. We fill the available states in a metal just as we filled the available states in an atom: we add electrons, one per quantum state, starting at the lowest energy and ending when we have added all the necessary electrons to the metal.

Let us first consider conditions at the absolute zero of temperature. This represents the lowest energy state of our sample, and we achieve it by placing the conduction electrons into the unfilled states that lie lowest in energy. This process is suggested by Fig. 49-1*b*, which shows the *probability function* $p(E)$. This function gives the probability of the state at the energy E to be occupied. At $T = 0$, all states below a certain energy are filled ($p = 1$) and all

---

*See *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles,* by Robert Eisberg and Robert Resnick (Wiley, 1985) 2nd ed., Section 11-11.

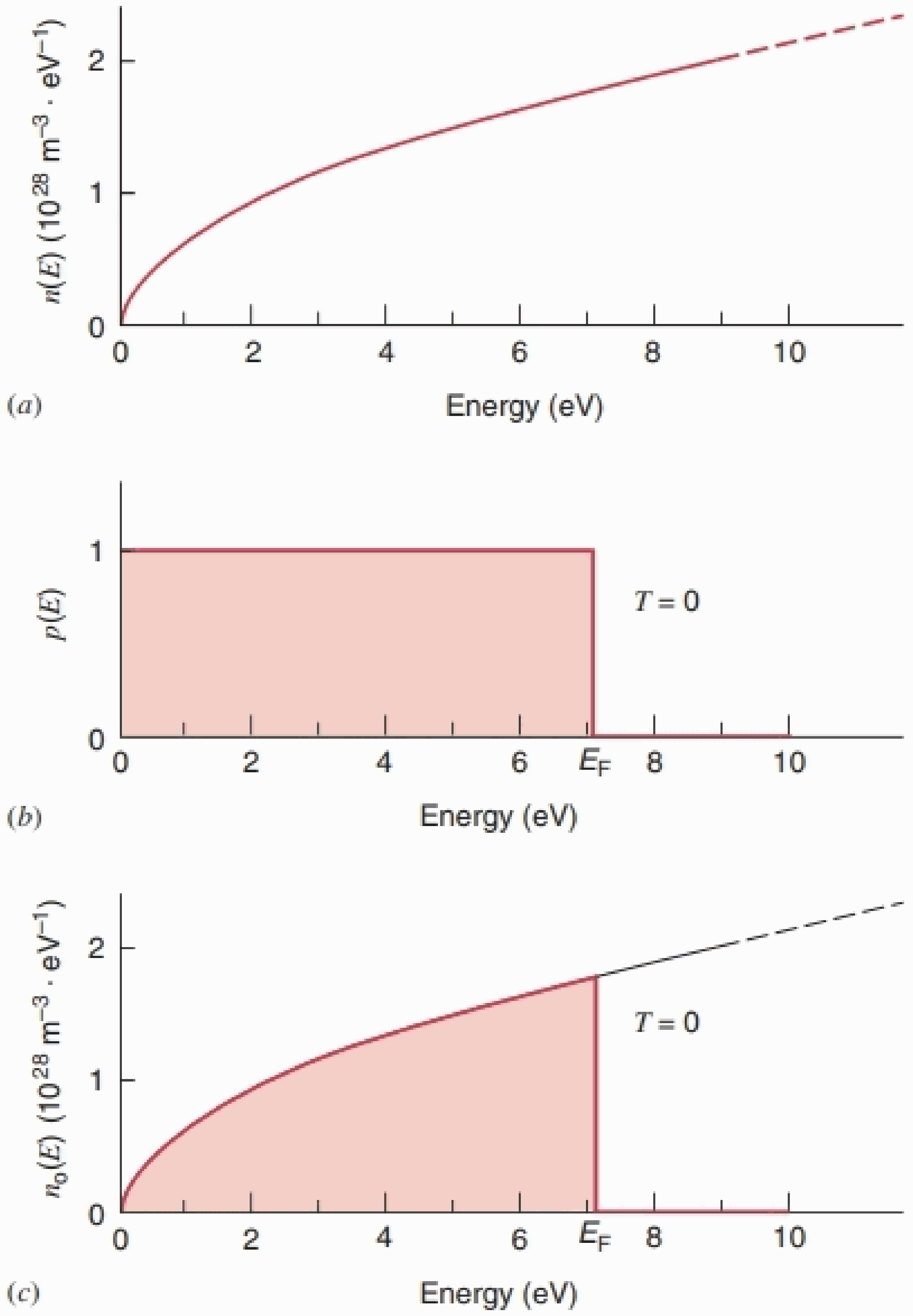


FIGURE 49-1. (*a*) The density of states $n(E)$ plotted as a function of the energy E. (*b*) The probability function $p(E)$ at $T = 0$. (*c*) The density of occupied states $n_o(E)$, equal to the product of $n(E)$ and $p(E)$. All states below E_F are occupied and all states above E_F are vacant.

states above that energy are vacant ($p = 0$). The highest occupied state under these conditions is called the *Fermi level,* and its energy, marked E_F in Fig. 49-1*b*, is called the *Fermi energy.* The Fermi energy for copper, for example, is 7.06 eV.

If we multiply the density $n(E)$ of *available* states by the probability $p(E)$ that those states are occupied, the result is the *density of occupied states,* $n_o(E)$, or

$$n_o(E) = n(E)p(E). \tag{49-3}$$

This quantity is plotted in Fig. 49-1*c*.

The shaded area in Fig. 49-1*c* represents the total number of occupied states (per unit volume). Finding this area and equating it to the density n of conduction electrons in the metal gives a means to find the Fermi energy. Integrating between the limits of $E = 0$ and $E = E_F$ to find the area, we obtain

$$n = \int_0^{E_F} n_o(E)dE, \tag{49-4}$$

or, carrying out the integral,

$$n = \frac{8\sqrt{2}\pi m^{3/2}}{h^3}\int_0^{E_F} E^{1/2}\,dE = \left(\frac{8\sqrt{2}\pi m^{3/2}}{h^3}\right)\left(\tfrac{2}{3}E_F^{3/2}\right).$$

Solving for E_F gives

$$E_F = \frac{h^2}{8m}\left(\frac{3n}{\pi}\right)^{2/3}. \tag{49-5}$$

A glance at Fig. 49-1*c* should be enough to shatter the popular misconception that all motion ceases at the absolute zero of temperature. We see that, entirely because of Pauli's exclusion principle, the electrons are stacked up in energy from zero to the Fermi energy. The *average* energy for the conditions of Fig. 49-1*c* turns out to be about 4.2 eV. By comparison, the average translational kinetic energy of a molecule of an ideal gas at room temperature is only 0.025 eV. The conduction electrons in a metal have plenty of energy at the absolute zero!

It seems clear that the molecules of a gas at ordinary temperatures and the conduction electrons in a metal behave in quite different ways. Formally, we say that the gas molecules obey the (classical) Maxwell–Boltzmann statistics and that the conduction electrons obey the (quantum) Fermi–Dirac statistics. The word "statistics" here refers to the formal rules for counting particles. In Maxwell–Boltzmann statistics, for example, we assume that we can tell identical particles apart, but in Fermi–Dirac statistics we assume that we cannot. Again, in Maxwell–Boltzmann statistics, Pauli's exclusion principle plays no role, but in Fermi–Dirac statistics its role, as we have seen, is vital.

What happens to the electron distribution of Fig. 49-1 as we raise the temperature? Only a small change occurs in the distribution, but that small change has important consequences. Figure 49-2 shows what the distributions of Fig. 49-1 would look like at $T = 1000$ K, a temperature at which a metal sample would glow brightly in a dark room.

The striking feature of Fig. 49-2 is how little it differs from Fig. 49-1, the distributions at absolute zero. At $T = 0$, the probability function $p(E)$ was strictly unity below E_F and strictly zero above E_F. As Fig. 49-2*b* shows, at $T = 1000$ K there is a small probability of having a few vacant states below E_F and a few occupied states above E_F. The density $n_o(E)$ of occupied states, again given by Eq. 49-3 as the product of $n(E)$ and $p(E)$, is shown in Fig. 49-2*c*. Because a few states above E_F are occupied, the average energy is a little larger than it was at the absolute zero but it is not much larger. This is again in striking contrast with the behavior of an ideal gas, for which the average kinetic energy of the molecules is proportional to the temperature.

By comparing conditions at $T = 0$ and at $T = 1000$ K, we see that all the "action" takes place for conduction electrons whose energies are close to the Fermi energy. The motion of most of the electrons remains unchanged as the temperature is raised, because their large store of energy is effectively locked up.

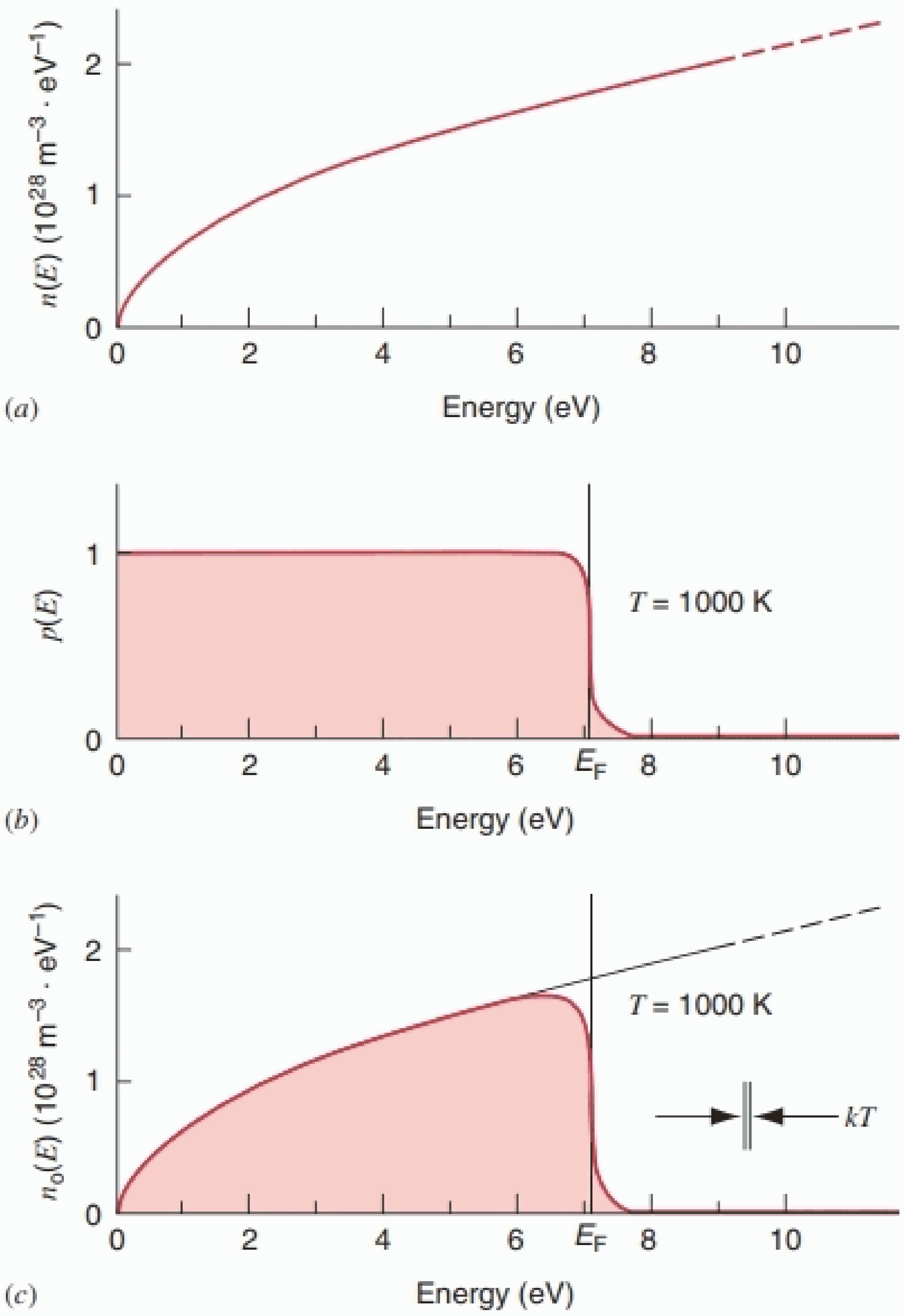


FIGURE 49-2. Same as Fig. 49-1, but for $T = 1000$ K. Note how little the plots differ from those of Fig. 49-1. (These plots are somewhat idealized in that they assume the electrons move in a region of uniform potential. Measured density of states plots in real metals do not have this simple shape.)

Let us see why this is so. Figure 49-2*c* displays the magnitude of kT, a measure of the energy available from thermal agitation; its value at 1000 K is only 0.086 eV. No electron can hope to have its energy changed by more than a few times this relatively small amount by thermal agitation alone. Because of the exclusion principle, only electrons whose energies are near the Fermi energy have vacant states close enough to them for such thermal transitions to occur. An electron with an energy of, say, 2 eV can neither gain nor lose energy because all states close enough to it in energy are already filled; it simply has nowhere to go. In analogy with waves on the ocean, thermal agitation of the electrons normally causes only ripples on the surface of the "Fermi sea"; the vast depths of that sea lie undisturbed.

The probability function $p(E)$ plotted in Figs. 49-1*b* and 49-2*b* is called the *Fermi–Dirac probability function* and can be shown to be

$$p(E) = \frac{1}{e^{(E-E_F)/kT} + 1}, \tag{49-6}$$

in which E_F is the Fermi energy, now defined (see Fig. 49-2*b*) as the energy corresponding to $p = \frac{1}{2}$.

Note that Eq. 49-6 yields the rectangular plot in Fig. 49-1*b* for $T = 0$. As $T \rightarrow 0$, the exponent $(E - E_F)/kT$ in Eq. 49-6 approaches $-\infty$ if $E < E_F$ and $+\infty$ if $E > E_F$. In the first case we have $p(E) = 1$ and in the second case $p(E) = 0$, just as required.

Equation 49-6 also shows us that the important quantity is not the energy E but rather $E - E_F$, the energy interval between E and the Fermi energy. We see further that, because of the exponential nature of the term in the denominator of Eq. 49-6, $p(E)$ is very sensitive to small changes in $E - E_F$. This confirms our assertion that electrons whose energies are close to the Fermi energy are the only ones that play an active role. As we shall see, the first question in dealing with the electrons in a solid—be it a conductor, a semiconductor, or an insulator—is likely to be: "On an energy scale, where is the Fermi level?"

SAMPLE PROBLEM 49-2. Calculate the Fermi energy for copper, given that the number of conduction electrons per unit volume (see Sample Problem 29-3) is $8.49 \times 10^{28}\ \mathrm{m^{-3}}$.

Solution From Eq. 49-5 we obtain

$$\begin{aligned} E_F &= \frac{h^2}{8m}\left(\frac{3n}{\pi}\right)^{2/3} \\ &= \frac{(6.63 \times 10^{-34}\ \mathrm{J\cdot s})^2}{(8)(9.11 \times 10^{-31}\ \mathrm{kg})}\left[\frac{(3)(8.49 \times 10^{28}\ \mathrm{m^{-3}})}{\pi}\right]^{2/3} \\ &= 1.13 \times 10^{-18}\ \mathrm{J} = 7.06\ \mathrm{eV}. \end{aligned}$$

SAMPLE PROBLEM 49-3. What is the probability of occupancy for a state whose energy is (*a*) 0.1 eV above the Fermi energy, (*b*) 0.1 eV below the Fermi energy, and (*c*) equal to the Fermi energy? Assume a temperature of 800 K.

Solution (*a*) The (dimensionless) exponent in Eq. 49-6 is

$$\frac{E - E_F}{kT} = \frac{0.1\ \mathrm{eV}}{(8.62 \times 10^{-5}\ \mathrm{eV/K})(800\ \mathrm{K})} = 1.45.$$

Inserting this exponent into Eq. 49-6 yields

$$p(E_+) = \frac{1}{e^{1.45} + 1} = 0.19.$$

Thus the occupancy probability for this state is 19%.

(*b*) For an energy 0.1 eV *below* the Fermi energy, the exponent in Eq. 49-6 has the same numerical value as above but is negative. Thus, from Eq. 49-6,

$$p(E_-) = \frac{1}{e^{-1.45} + 1} = 0.81.$$

The occupancy probability for this state is 81%.

(*c*) For $E = E_F$ the exponent in Eq. 49-6 is zero and that equation becomes

$$p(E_F) = \frac{1}{e^0 + 1} = \frac{1}{1 + 1} = 0.50.$$

Note that this result does not depend on the temperature. Note also that none of these three results depends on the actual value of the Fermi energy, only on the energy interval between the Fermi energy and the energy of the state in question.

SAMPLE PROBLEM 49-4. (*a*) For copper at 1000 K, find the energy at which the probability $p(E)$ that a conduction electron state will be occupied is 90%. (Assume that the conduction electrons in copper behave like a free-electron gas, with a Fermi energy of 7.06 eV.) (*b*) For this energy, what is $n(E)$, the distribution in energy of the available states? (*c*) For this same energy, what is $n_o(E)$, the distribution in energy of the occupied states?

Solution (*a*) Substitution into Eq. 49-6 yields

$$p(E) = \frac{1}{e^{\Delta E/kT} + 1} = 0.9,$$

in which $\Delta E = E - E_F$. A little algebra leads to $\Delta E/kT = -2.20$ so that

$$\begin{aligned} \Delta E &= -2.20kT = -(2.20)(8.62 \times 10^{-5}\ \mathrm{eV/K})(1000\ \mathrm{K}) \\ &= -0.19\ \mathrm{eV}. \end{aligned}$$

For copper, assuming that $E_F = 7.06$ eV, we have

$$E = E_F + \Delta E = 7.06\ \mathrm{eV} - 0.19\ \mathrm{eV} = 6.87\ \mathrm{eV}.$$

(*b*) Carrying out a calculation just like that of Sample Problem 49-1 for $E = 6.87$ eV yields $n(E) = 1.78 \times 10^{28}\ \mathrm{m^{-3}\,eV^{-1}}$.
(*c*) From Eq. 49-3 we have, again for $E = 6.87$ eV,

$$\begin{aligned} n_o(E) = n(E)p(E) &= (1.78 \times 10^{28}\ \mathrm{m^{-3}\,eV^{-1}})(0.90) \\ &= 1.60 \times 10^{28}\ \mathrm{m^{-3}\,eV^{-1}}. \end{aligned}$$

49-4 ELECTRICAL CONDUCTION IN METALS

Figure 49-3*a* represents the Fermi distribution of *velocities* in a metal. The *Fermi speed* v_F is the speed of an electron whose kinetic energy equals E_F, the Fermi energy. With no applied electric field, electrons have speeds ranging from 0 to approximately v_F, corresponding to energies ranging from 0 to approximately E_F. The distribution in Fig. 49-3 represents a typical *velocity* component, rather than the speed. This illustrates that there are equal numbers of electrons moving in opposite directions, so that the net current is zero in the absence of an electric field.

When an electric field is applied, the electrons are accelerated by the field and acquire a small increase in velocity in a direction opposite to the field. (Because electrons are negatively charged, the force on an electron is $\vec{\mathbf{F}} = -e\vec{\mathbf{E}}$, which is in a direction opposite to $\vec{\mathbf{E}}$.) The entire velocity distribution in the presence of a field is shifted slightly to the right as in Fig. 49-3*b*. However, most of the electrons still add pairwise to zero velocity and do not contribute to the conduction.

The electrons that contribute to the conduction are those in a small group of velocities near v_F. The electric field causes states having velocities just below v_F in the direction of $\vec{\mathbf{E}}$ to become unoccupied, while states having velocities just above v_F in a direction opposite to $\vec{\mathbf{E}}$ become occupied. You can see from Fig. 49-3 why the drift velocity (the average velocity of all the electrons) is much smaller than v_F, because in the averaging process many positive and negative velocities will cancel one another. The drift speed is determined primarily by the small number of electrons moving from states below speed v_F to states above speed v_F under the action of the electric field.

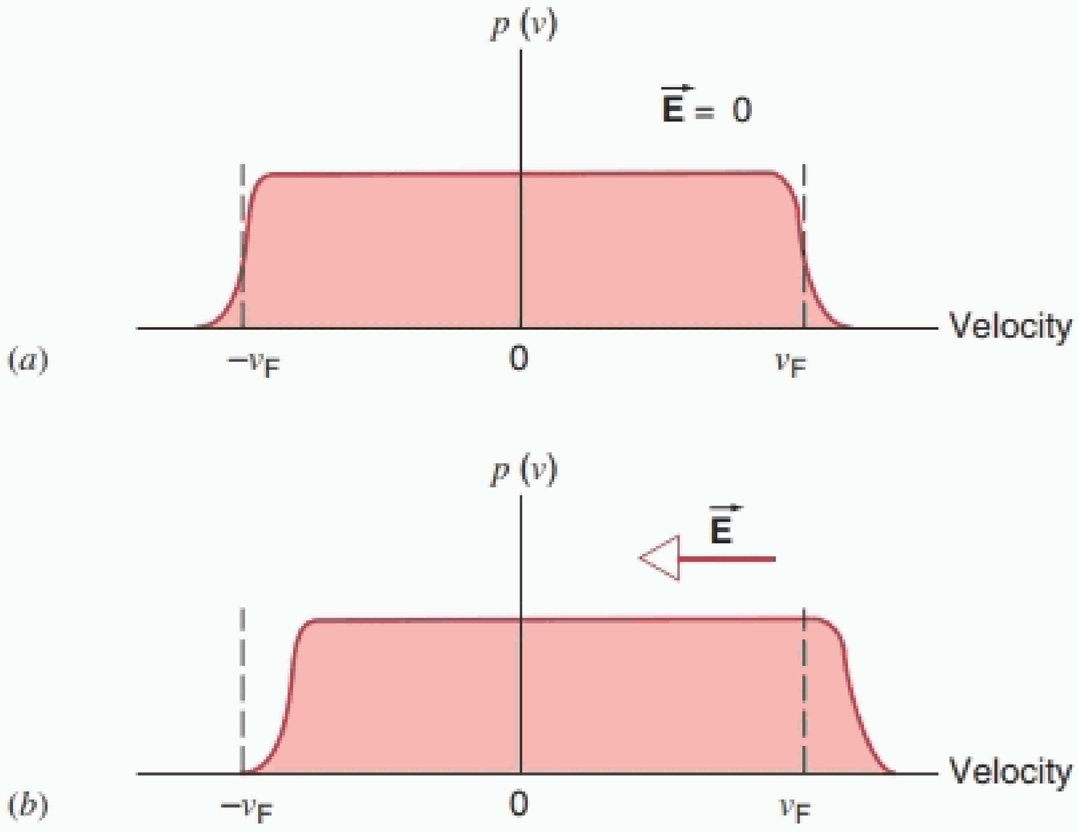


FIGURE 49-3. The Fermi distribution of velocities. (*a*) With no electric field states up to the Fermi speed v_F are filled. (*b*) When an electric field $\vec{\mathbf{E}}$ is applied in the direction shown, the distribution shifts to the right as the electrons are accelerated by the field.

The resistivity of the metal to the flow of these electrons is determined by collisions made by the electrons with the ion cores of the lattice. In Sample Problem 49-5 we show that, for copper at room temperature, the Fermi speed, which is the average speed of the conduction electrons between collisions, is 1.6×10^6 m/s, a significant fraction of the speed of light. The mean time between collisions is 2.5×10^{-14} s and the mean free path is 41 nm, which is about 150 nearest-neighbor distances in the copper lattice.

You may be surprised that, at room temperature, a conduction electron can move so far through a copper lattice without hitting an ion core. At lower temperatures—where the resistivity is lower—it can move even much further. It is, in fact, a perhaps unexpected prediction of wave mechanics that a perfectly periodic lattice at the absolute zero of temperature would be totally transparent to conduction electrons. There would never be any collisions!

There are, however, no perfectly periodic lattices. Vacant lattice sites and impurity atoms are always present, no matter how hard we try to eliminate them. Furthermore, at temperatures above the absolute zero the lattice is vibrating, and these motions also spoil the periodicity of the lattice. At room temperature the "collisions" of which we have spoken are largely interactions between the conduction electrons and the vibrations of the lattice.

SAMPLE PROBLEM 49-5. Take the Fermi energy of copper to be 7.06 eV. (*a*) What is the speed of a conduction electron with this kinetic energy? (*b*) The resistivity ρ of copper at room temperature is $1.69 \times 10^{-8}\ \Omega \cdot \text{m}$ and n, the number of electrons per unit volume, is $8.49 \times 10^{28}\ \text{m}^{-3}$. What is the average time τ between collisions? (*c*) What mean free path λ may be calculated from the results of (*a*) and (*b*)?

Solution (*a*) Throughout this section we have assumed that the conduction electrons are moving in a region in which their potential energy is zero. Thus their total energy E is all kinetic and we can write, if $E = E_\text{F}$,

$$E_\text{F} = \tfrac{1}{2} m v_\text{F}^2,$$

in which v_F is the Fermi speed. Solving for v_F yields

$$v_\text{F} = \left(\frac{2E_\text{F}}{m}\right)^{1/2} = \left[\frac{(2)(7.06\ \text{eV})(1.60 \times 10^{-19}\ \text{J/eV})}{9.11 \times 10^{-31}\ \text{kg}}\right]^{1/2}$$
$$= 1.57 \times 10^6\ \text{m/s}.$$

You should not confuse this speed with the *drift speed* of the conduction electrons, which is of the order of 10^{-4} m/s and is thus smaller by about a factor of 10^{10}. As we explained in Section 29-5, the drift speed is the average speed at which electrons actually drift through a conductor when an electric field is applied; the Fermi speed is their average speed between collisions.

(*b*) Solving Eq. 49-1 for τ yields

$$\tau = \frac{m}{ne^2\rho}$$
$$= \frac{9.11 \times 10^{-31}\ \text{kg}}{(8.49 \times 10^{28}\ \text{m}^{-3})(1.60 \times 10^{-19}\ \text{C})^2(1.69 \times 10^{-8}\ \Omega \cdot \text{m})}$$
$$= 2.48 \times 10^{-14}\ \text{s}.$$

(*c*) To find the mean free path, we have

$$\lambda = v_\text{F}\tau = (1.57 \times 10^6\ \text{m/s})(2.48 \times 10^{-14}\ \text{s})$$
$$= 3.9 \times 10^{-8}\ \text{m} = 39\ \text{nm}.$$

In the copper lattice the centers of neighboring ion cores are 0.26 nm apart. Thus a typical conduction electron can move a substantial distance through a copper lattice at room temperature without making a collision.

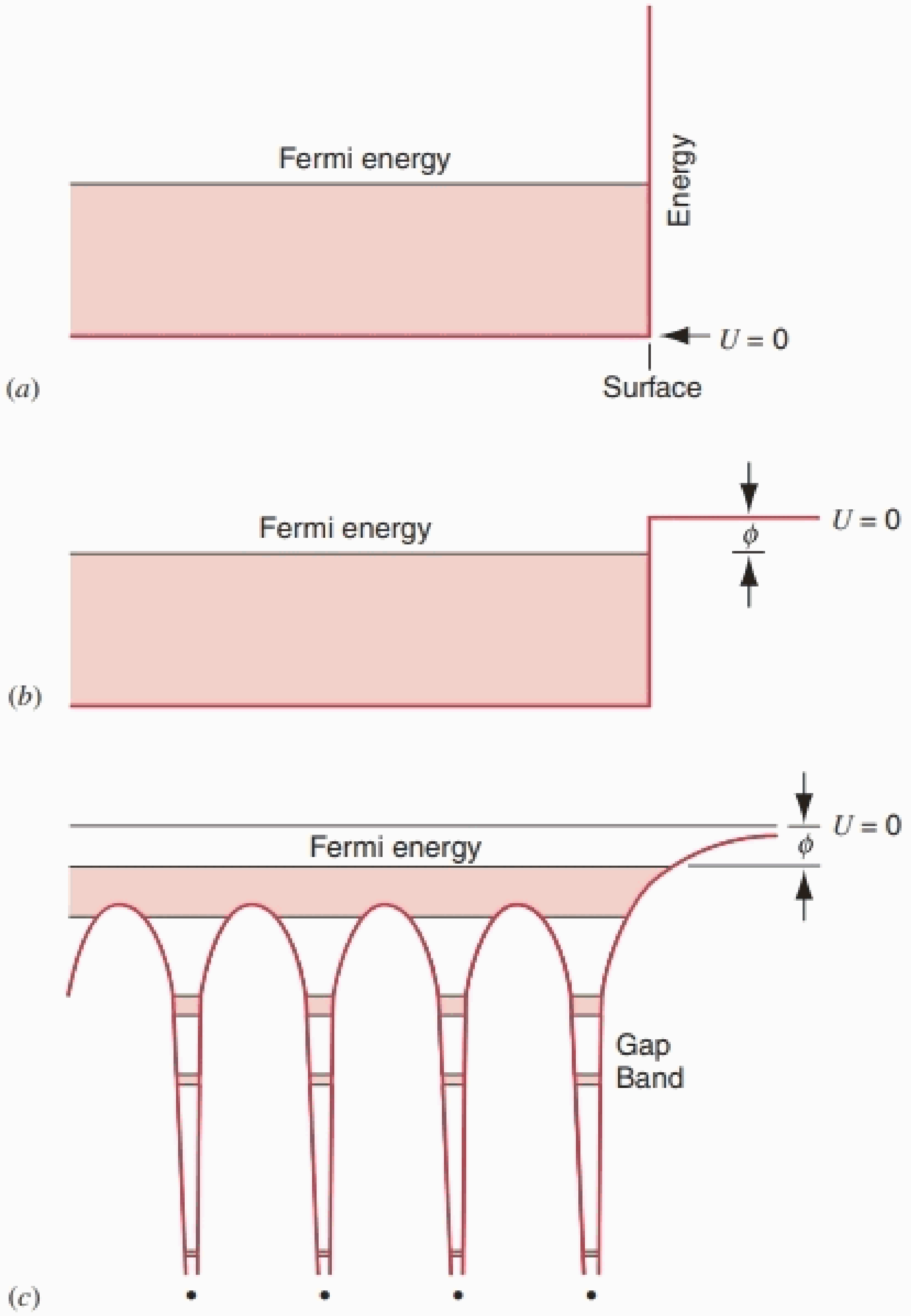


FIGURE 49-4. (*a*) The potential energy variation assumed for a metal in the free-electron gas model. (*b*) A more realistic variation, showing a finite change in potential energy at the surface of the sample. (*c*) A still more realistic variation, taking the lattice of ion cores into account. This curve is a one-dimensional cut along a line of ion cores (shown as dots at the bottom of the figure). The shaded regions are the energy bands permitted for the electrons. Electrons are not permitted to have energies corresponding to the gaps.

49-5 BANDS AND GAPS

Figure 49-4*a* suggests the potential energy variation that we have been using to describe a conduction electron in a metal. The potential energy is zero inside the metal, and it rises to infinity at the surface. However, there are problems with this model. For example, it tells us that, because of the infinite potential barrier, an electron could never escape from inside the sample through its surface. We know that this is not true, because electrons can be "boiled out" of a metal by raising its temperature, as in the heated filament of a vacuum tube (thermionic emission). They can also be "kicked out" if we shine light of high enough frequency on the metal surface (photoelectric effect).

Figure 49-4*b* shows that we can take care of this difficulty easily enough by making the potential energy at the surface finite. We made the same realistic adjustment (see Fig. 47-6) for the electron trapped in an atom-sized, one-dimensional well. The quantity ϕ in Fig. 49-4*b* is the *work function* of the metal, defined as the least amount of energy that must be supplied to an electron to remove it from the sample.

Figure 49-4*b* has been drawn to bring our energy scale into agreement with that used for the hydrogen atom. That is, we have chosen the $E = 0$ configuration to represent an electron at rest far outside the sample. It is possible to make this change because the potential energy always contains an arbitrary additive constant and we are more concerned, in any case, with changes in the total energy E than with E itself. On our new energy scale, the total energies of electrons trapped in the sample are negative, just as they are for the hydrogen atom.

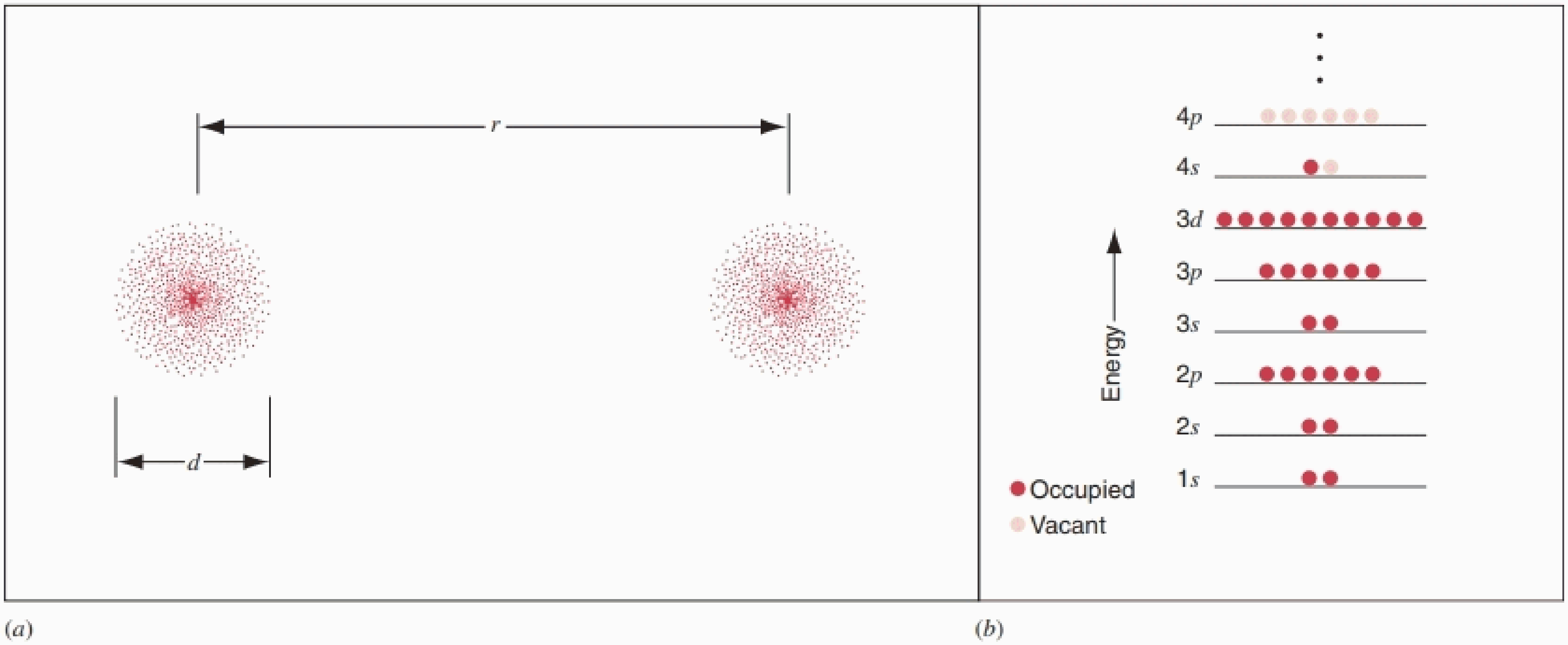


FIGURE 49-5. (*a*) Two neutral copper atoms of diameter *d*, separated by a distance *r*, with $r \gg d$. (*b*) The atoms are independent systems, and in its ground state each has the same quantum number assignments for its electrons, as shown. The energy scale is symbolic only.

By far the largest difficulty remaining in Fig. 49-4*b* is that it assumes that the potential energy of a conduction electron is constant throughout the volume of the sample. This ignores the fact that the conduction electrons move in an array of positively charged ion cores. It is rather remarkable, in fact, that we have been able to learn as much as we have about the resistivity of metals without taking into account the potential energy variations caused by the ion cores of the lattice. We have not, however, been able to answer such questions as: "Why is copper a conductor and diamond a nonconductor?" If we take the lattice periodicity into account, we shall be able to answer this question and to go far beyond.

Figure 49-4*c* shows a potential energy curve that takes the ion cores into account. Substituting this potential energy (or some approximation to it) into the Schrödinger equation brings out an interesting new phenomenon. As Fig. 49-4*c* shows, *the allowed states are now grouped into bands, with energy gaps between them in which no states exist.* Note that electrons just below the Fermi level are free to move throughout the lattice, but electrons with lower energies, the *core electrons,* are not. Let us see if we can understand these bands and gaps in physical terms.

The distance between nearest neighbors in a copper lattice is 0.26 nm. Consider, however, two copper atoms separated by a much greater distance—say, 50 nm—so that we may describe them as "isolated"; see Fig. 49-5*a*. In each atom the 29 electrons are assigned to the levels shown in Fig. 49-5*b*.

Now let us bring the two atoms closer together so that an outer electron in either atom can be influenced, however slightly, by forces exerted on it by the other atom. In the language of quantum mechanics we say that their wave functions begin to overlap. We state without proof that the two overlapping wave functions can be combined in two independent ways, describing two states having (slightly) different energies, as shown in the second column of Fig. 49-6. Because the overlap is greater for the outer electrons, the energy splitting will be greater for them than for the inner electrons. If we have three interacting atoms, the individual levels of the isolated atoms become three closely spaced levels.

By extension, if we bring *N* copper atoms together to form a copper lattice, each level of the isolated atom becomes *N* closely spaced levels of the solid. Thus the 1*s* level of the atom becomes the 1*s* band of the solid, and so on. Figure 49-6 suggests the process.

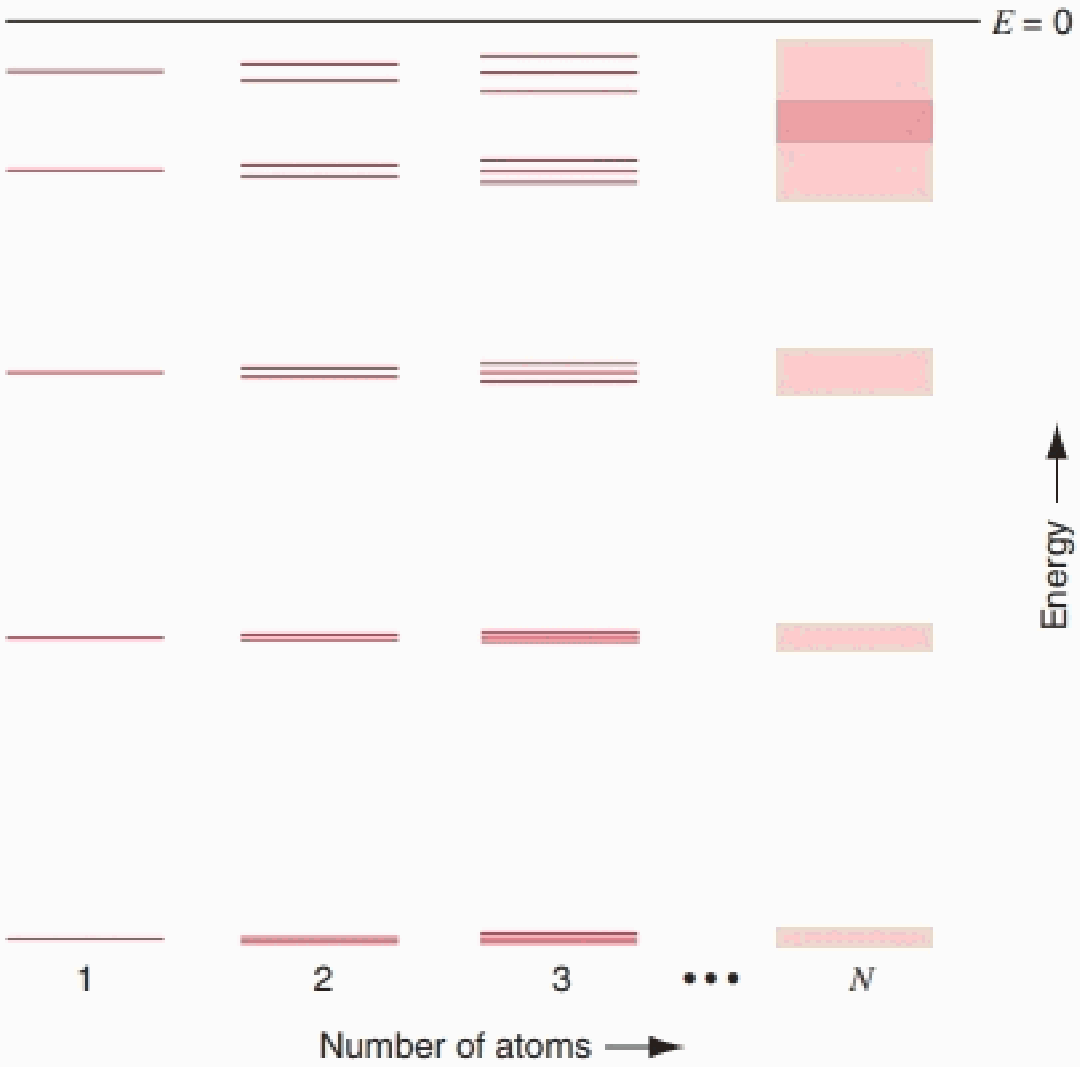


FIGURE 49-6. As atoms are brought together to form a lattice, the levels of the isolated atoms split, eventually forming bands of closely lying levels. For the case shown, the upper bands overlap in energy.

From this point of view the forbidden gaps are not so hard to understand, being familiar from the level structure of the isolated atom. Indeed, we can say that Niels Bohr, even before wave mechanics, "invented" energy gaps when he said, in effect: "I assume that atoms can exist without radiating in a discrete set of stationary states of definite energy, *states of intermediate energy being forbidden.*"

49-6 CONDUCTORS, INSULATORS, AND SEMICONDUCTORS

Figure 49-7*a* represents the band structure of a conductor, such as copper. Its central feature is that the highest-energy populated band is only partially filled. There are vacant states above the Fermi level so that, if you apply an electric field $\vec{\mathbf{E}}$, every electron in this band is able to increase its momentum in the $-\vec{\mathbf{E}}$ direction, and there will be a current. The lower energy bands are completely filled and cannot contribute to the conduction process, all velocities adding pairwise to zero.

Figure 49-7*b* represents an insulator. Its central feature is that the highest-energy populated band is completely filled, and the forbidden energy gap lying immediately above it, marked E_g in the figure, is substantial. By "substantial" we mean that $E_g \gg kT$, so the probability that an electron will be lifted, by thermal agitation, into the empty band that lies above the gap is negligible. If you set up an electric field within an insulator, there is no way for any of the electrons to respond to it, so there will be no current.

Carbon in its diamond form is an excellent insulator because its energy gap is 5.5 eV, more than 200 times the value of kT at room temperature.

Figure 49-7*c* represents a semiconductor. It differs from an insulator in that its energy gap is small enough so that thermal excitation of electrons across it can occur to some extent at room temperature. This puts electrons into the (nearly empty) band labeled *conduction band* in the figure and leaves an equal number of vacant states, or *holes,* in the (nearly filled) *valence band.* In a band that is nearly full, it turns out to be more convenient to analyze its contribution to the electrical conduction in terms of the motion of holes, which behave like *positively* charged particles.

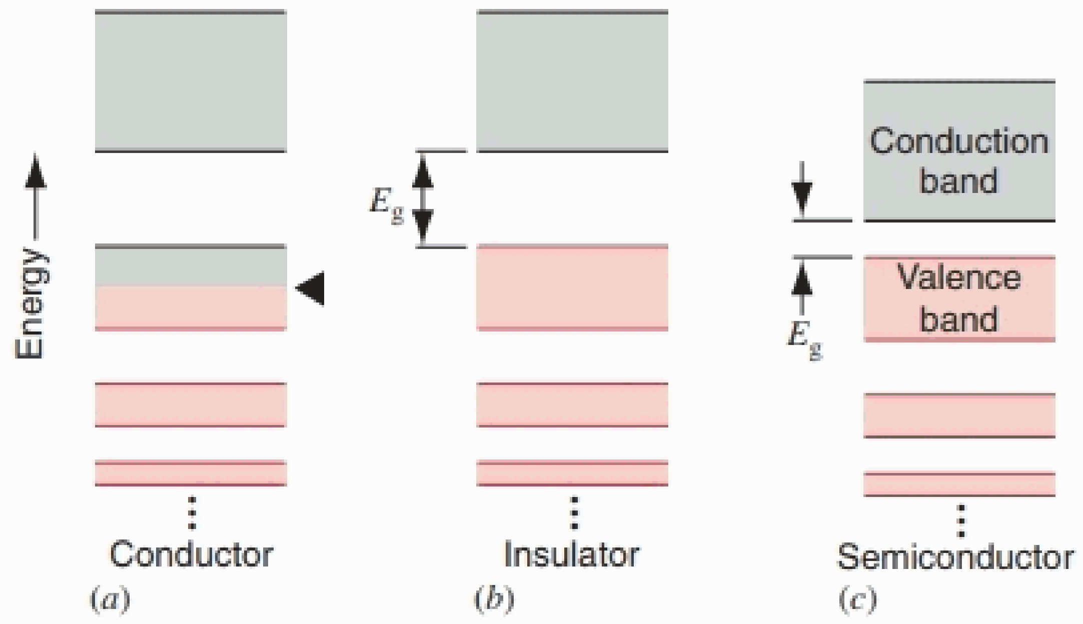


FIGURE 49-7. An idealized representation of the energy bands for (*a*) a conductor, (*b*) an insulator, and (*c*) a semiconductor. Filled bands are shown in colored shading, and empty bands in gray shading. The black triangle marks the Fermi level for the conductor.

Silicon is a typical semiconductor. It has the same crystal structure as diamond but its gap width (= 1.1 eV) is considerably smaller. At the absolute zero of temperature, where thermal agitation is absent, all semiconductors are insulators. At any higher temperature the probability that an electron will be raised across the gap is very sensitive to the gap width. Thus the distinction between an insulator and a semiconductor is somewhat arbitrary. There is no hesitation, however, in calling diamond (E_g = 5.5 eV) an insulator and silicon (E_g = 1.1 eV) a semiconductor.

SAMPLE PROBLEM 49-6. What is the probability at room temperature that a state at the bottom of the conduction band is occupied in diamond and in silicon? Take the Fermi energy to be at the middle of the gap between the conduction and valence bands.

Solution For a state at the bottom of the conduction band, the energy difference $E - E_F$ is $0.5E_g$, if the Fermi energy is at the middle of the gap. At room temperature (300 K), $kT = 0.026$ eV. We therefore have $E - E_F \gg kT$, and so we can approximate the Fermi–Dirac probability function (Eq. 49-6) as

$$p(E) = \frac{1}{e^{(E-E_F)/kT} + 1} \approx e^{-(E-E_F)/kT} = e^{-E_g/2kT}.$$

For diamond, $E_g = 5.5$ eV and so

$$p(E) = e^{-(5.5\ \mathrm{eV})/2(0.026\ \mathrm{eV})} = 1.2 \times 10^{-46}.$$

For silicon, $E_g = 1.1$ eV and so

$$p(E) = e^{-(1.1\ \mathrm{eV})/2(0.026\ \mathrm{eV})} = 6.5 \times 10^{-10}.$$

In a cubic centimeter of material, containing roughly 10^{23} atoms, there will be a negligible probability of finding even one electron in the conduction band of diamond, while there may be roughly 10^{13} electrons in the conduction band (and an equal number of holes in the valence band) available for electrical conduction in silicon. This calculation illustrates the extreme difference in conductivity that results from small variations in the gap energy, and it clearly shows the distinction between insulators and semiconductors. In a cubic centimeter of a conductor, on the other hand, there might be 10^{23} electrons available for electrical conduction.

Semiconductors

In Sample Problem 49-6, we compared a property of a semiconductor with that of an insulator. Table 49-1 compares some properties of a typical semiconductor (silicon) and a typical conductor (copper). Let us now discuss these properties in more detail.

1. *The density of charge carriers, n.* Copper has many more charge carriers than does silicon, by a factor of about 10^{13}. For copper the carriers are the conduction electrons,

TABLE 49-1 Some Electrical Properties of Copper and Silicon[a]

	Copper	*Silicon*
Type of material	Conductor	Semiconductor
Density of charge carriers[b] n (m^{-3})	9×10^{28}	1×10^{16}
Resistivity ρ ($\Omega \cdot \text{m}$)	2×10^{-8}	3×10^{3}
Temperature coefficient of resistivity α (K^{-1})	$+4 \times 10^{-3}$	-70×10^{-3}

[a] All values refer to room temperature.

[b] Includes, for silicon, both electrons and holes.

about one per atom. Figure 49-7*c* shows that, at the absolute zero of temperature, silicon would have no charge carriers at all. At room temperature, to which Table 49-1 refers, charge carriers arise only because, at thermal equilibrium, thermal agitation has caused a certain (very small) number of electrons to be raised to the conduction band, leaving an equal number of vacant states (holes) in the valence band.

The holes in the valence band of a semiconductor also serve effectively as charge carriers because they permit a certain freedom of movement to the electrons in that band. If an electric field is set up in a semiconductor, the electrons in the valence band, being negatively charged, drift in the direction of $-\vec{\mathbf{E}}$. The holes drift in the direction of the field and behave like particles carrying a charge $+e$, which is exactly how we shall regard them. Conduction by holes is an important characteristic of semiconductors.

2. *The resistivity, ρ.* At room temperature the resistivity of silicon is considerably higher than that of copper, by a factor of about 10^{11}. For both elements, the resistivity is determined by Eq. 49-1. As that equation shows, the resistivity increases as n, the density of charge carriers, decreases. The vast difference in resistivity between copper and silicon can be accounted for by the vast difference in n. (The mean collision time τ will also be different for copper and for silicon, but the effect of this on the resistivity is overwhelmed by the enormous difference in the density of charge carriers.)

For completeness, we mention that the resistivity of a good insulator (fused quartz or diamond, for example) may be as high as $10^{20}\ \Omega \cdot \text{m}$, about 10^{28} times higher than that of copper at room temperature. Few physical properties have as wide a range of measurable values as the electrical resistivity.

3. *The temperature coefficient of resistivity, α.* This quantity (see Eq. 29-18) is the fractional change in the resistivity ρ per unit change of temperature, or

$$\alpha = \frac{1}{\rho}\frac{d\rho}{dT}.$$

The resistivity of copper and other metals *increases* with temperature ($d\rho/dT > 0$). This happens because collisions occur more frequently as the temperature is increased, thus reducing τ in Eq. 49-1. For metals, the density of charge carriers n in that equation is independent of temperature.

On the other hand, the resistivity of silicon (and other semiconductors) *decreases* with temperature ($d\rho/dT < 0$). This happens because the density of charge carriers n in Eq. 49-1 increases rapidly with temperature. The decrease in τ mentioned above for metals also occurs for semiconductors, but its effect on the resistivity is overwhelmed by the very rapid increase of the density of charge carriers.

In the laboratory, you can identify a semiconductor by its large resistivity ρ and, especially, by its large—and negative—temperature coefficient of resistivity α, with both quantities compared to values for a typical metal.

49-7 DOPED SEMICONDUCTORS

The performance of semiconductors can be substantially changed by deliberately introducing a small number of suitable replacement atoms as impurities into the semiconductor lattice, a process called *doping.* We describe the semiconductor that results as *extrinsic,* to distinguish it from the pure undoped or *intrinsic* material. Essentially all semiconducting devices today are based on extrinsic material.

Figure 49-8*a* is a two-dimensional representation of a lattice of pure silicon. Each silicon atom has four valence electrons and forms a two-electron bond with each of its four nearest neighbors, the electrons involved in the bonding making up the valence band of the sample.

In Fig. 49-8*b* one of the silicon atoms has been replaced by an atom of phosphorus, which has five valence electrons. Four of these electrons form bonds with the four neighboring silicon atoms, but the fifth electron is loosely bound to the phosphorus ion core, as Fig. 49-8*b* suggests. It is far easier for *this* electron to be thermally excited into the conduction band than it is for one of the silicon valence electrons to be so excited.

The phosphorus atom is called a *donor* atom because it readily *donates* an electron to the conduction band. The "extra" electron in Fig. 49-8*b* can be said to lie in a localized donor level, as Fig. 49-9*a* shows. This level is separated from the bottom of the conduction band by an energy gap E_d, where usually $E_\text{d} \ll E_\text{g}$. Adding donor atoms can greatly increase the density of electrons in the conduction band.

Semiconductors doped with donor atoms are called *n-type* semiconductors, where the "*n*" stands for "negative" because the negative charge carriers (electrons) greatly outnumber the positive charge carriers (holes). In *n*-type semiconductors, the electrons in the conduction band are called the *majority carriers,* whereas the holes in the valence band are called the *minority carriers.*

Figure 49-8*c* shows a silicon lattice in which a silicon atom has been replaced by an aluminum atom, which has three valence electrons. In this case there is a "missing" electron, and it is easy for the aluminum ion core to "steal"

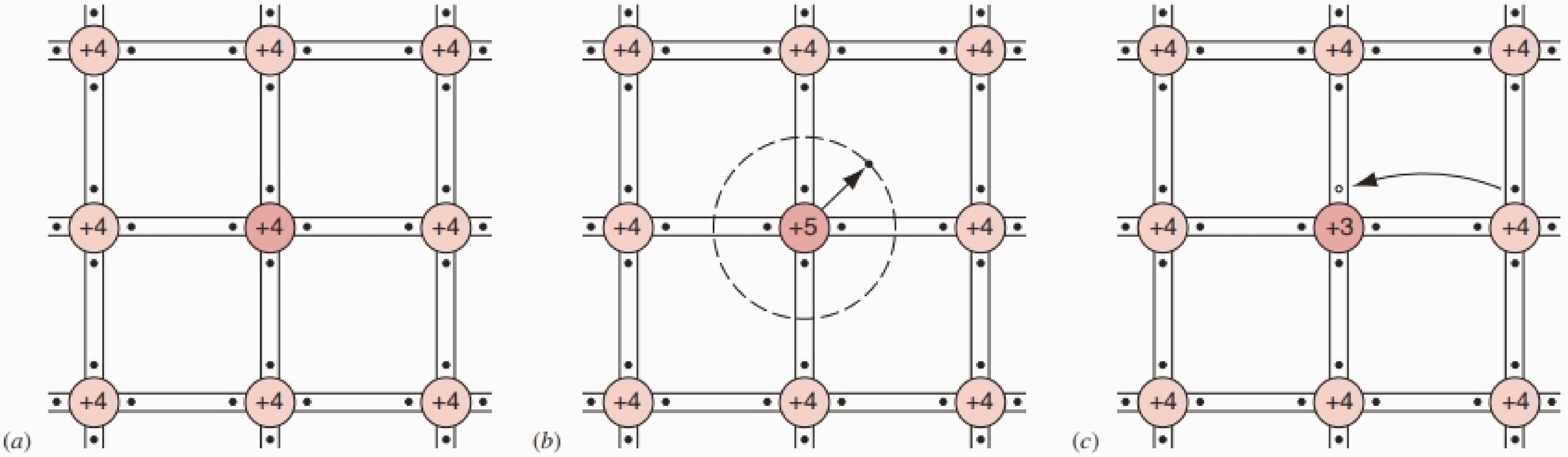


FIGURE 49-8. (*a*) A two-dimensional representation of the silicon lattice. Each silicon ion (core charge = +4*e*) is bonded to each of its four nearest neighbors by a shared two-electron bond. The dots show those valence electrons. (*b*) A phosphorus atom (valence = 5) is substituted for a silicon atom, creating a donor site. (*c*) An aluminum atom (valence = 3) is substituted for a silicon atom, creating an acceptor site.

a valence electron from a nearby silicon atom, thus creating a hole in the valence band.

The aluminum atom is called an *acceptor* atom because it so readily *accepts* an electron from the valence band. The electron so accepted moves into a localized acceptor level, as Fig. 49-9*b* shows. This level is separated from the top of the valence band by an energy gap E_a, where $E_a \ll E_g$. Adding acceptor atoms can greatly increase the number of holes in the valence band.

Semiconductors doped with acceptor atoms are called *p-type* semiconductors, where the "*p*" stands for "positive" because the positive charge carriers (holes) greatly outnumber the negative carriers (electrons). In *p*-type semiconductors the majority carriers are the holes in the valence band and the minority carriers are the electrons in the conduction band.

Table 49-2 summarizes the properties of a typical *n*-type and a typical *p*-type semiconductor. Note particularly that the donor and acceptor ion cores, although they are charged, are not charge carriers because, at normal temperatures, they remain fixed in their lattice sites.

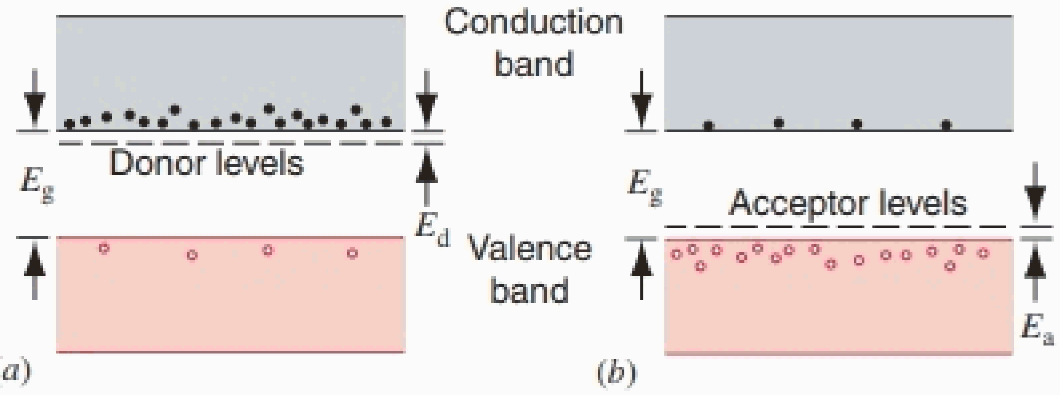


FIGURE 49-9. (*a*) An *n*-type semiconductor, showing donor levels that have contributed electrons (majority carriers) to the conduction band. The small number of holes (minority carriers) in the valence band is also shown. (*b*) A *p*-type semiconductor, showing acceptor levels that have contributed holes (majority carriers) to the valence band. The small number of electrons (minority carriers) in the conduction band is also shown.

TABLE 49-2 **Two Typical Extrinsic Semiconductors**

	n-Type	*p-Type*
Matrix material	Silicon	Silicon
Dopant	Phosphorus	Aluminum
Type of dopant	Donor	Acceptor
Dopant valence	5 (= 4 + 1)	3 (= 4 − 1)
Dopant energy gap	0.045 eV	0.057 eV
Majority carriers	Electrons	Holes
Minority carriers	Holes	Electrons
Dopant ion core charge	$+e$	$-e$

SAMPLE PROBLEM 49-7. The density of conduction electrons in pure silicon at room temperature is about $10^{16}\ \mathrm{m^{-3}}$. Assume that, by doping the lattice with phosphorus, you want to increase this number by a factor of 10^6. What fraction of the silicon atoms must you replace by phosphorus atoms? (Assume that, at room temperature, the thermal agitation is effective enough so that essentially every phosphorus atom donates its "extra" electron to the conduction band.)

Solution The density n_P of the phosphorus atoms must be about $(10^{16}\ \mathrm{m^{-3}})(10^6)$ or $10^{22}\ \mathrm{m^{-3}}$. The density of silicon atoms in a pure silicon lattice may be found from

$$n_{\mathrm{Si}} = \frac{N_A d}{M},$$

in which N_A is the Avogadro constant, d (= 2330 kg/m³) is the density of silicon, and M (= 28.1 g/mol) is the molar mass of silicon. Substituting yields

$$n_{\mathrm{Si}} = \frac{(6.02 \times 10^{23}\ \mathrm{mol^{-1}})(2330\ \mathrm{kg/m^3})}{0.0281\ \mathrm{kg/mol}} = 5 \times 10^{28}\ \mathrm{m^{-3}}.$$

The ratio of these two number densities is the quantity we are looking for. Thus

$$\frac{n_{\mathrm{Si}}}{n_{\mathrm{P}}} = \frac{5 \times 10^{28}\ \mathrm{m^{-3}}}{10^{22}\ \mathrm{m^{-3}}} = 5 \times 10^6.$$

We see that if only *one silicon atom in five million* is replaced by a phosphorus atom, the number of electrons in the conduction band will be increased by a factor of 10^6.

How can such a tiny admixture of phosphorus atoms have such a big effect? The answer is that, for pure silicon at room temperature, there were not many conduction electrons there to start with! The density of conduction electrons was $10^{16}\ \mathrm{m}^{-3}$ before doping and $10^{22}\ \mathrm{m}^{-3}$ after doping. For copper, however, the conduction electron density (see Table 49-1) is about $10^{29}\ \mathrm{m}^{-3}$. Thus, even *after* doping, the conduction electron density of silicon remains much less than that of a typical conductor such as copper.

SAMPLE PROBLEM 49-8. Assume that the "extra" electron in a phosphorus donor atom moves in a Bohr orbit around the central phosphorus ion core, as in Fig. 49-8*b*. Calculate (*a*) the binding energy and (*b*) the orbit radius for this electron.

Solution (*a*) The Bohr theory expression for the binding energy E_b of the $n = 1$ state is (see Eq. 47-23)

$$E_b = -E_1 = \frac{mZ^2e^4}{8\epsilon_0^2h^2}. \tag{49-7}$$

Here we put $Z = 1$ because the orbiting electron "sees" a net central charge of $+e$.

We derived the Bohr energy by considering a hydrogen-like atom, its orbiting electron moving in a vacuum. In this case, however, the electron moves through a silicon lattice. One effect of this is to reduce the electrostatic force by a factor of κ_e, the dielectric constant of silicon. To realize this force reduction quantitatively, we must replace ϵ_0 in Coulomb's law by $\kappa_e\epsilon_0$. Making the same replacement in Eq. 49-7 leads to

$$E_b = \frac{1}{\kappa_e^2}\left(\frac{me^4}{8\epsilon_0^2h^2}\right),$$

in which the factor in parentheses is just 13.6 eV, the binding energy of the hydrogen atom. For silicon we have $\kappa_e = 12$, so that

$$E_b = \frac{13.6\ \text{eV}}{(12)^2} = 0.094\ \text{eV}.$$

This result is in rough order-of-magnitude agreement with the value of 0.045 eV listed in Table 49-2.

(*b*) The orbit radius follows from Eq. 47-22. Substituting as in part (*a*) leads to

$$r = \kappa_e\left(\frac{\epsilon_0h^2}{\pi me^2}\right). \tag{49-8}$$

The factor in parentheses is just the Bohr radius (=52.9 pm). Thus

$$r = (12)(52.9\ \text{pm}) = 630\ \text{pm}.$$

This is comparable to the atomic spacing in the silicon lattice (540 pm).

We should also replace the electron mass m in Eqs. 49-7 and 49-8 by an effective electron mass m_{eff}, to take partially into account the periodic nature of the silicon lattice potential. Doing so reduces the estimated binding energy and increases the estimated orbit radius, both changes being in the direction of improving the agreement with experiment.

49-8 THE *pn* JUNCTION

In the next few sections we describe some commonly used semiconducting devices, such as diode rectifiers, light-emitting diodes, and transistors. There is no end to the number of such devices that we could have chosen to describe. With today's technology, in fact, it is possible to tailor-make complex semiconducting devices to meet specific needs.

Essentially all semiconducting devices involve one or more *pn junctions*. Consider a hypothetical plane through a rod of a pure crystalline semiconducting material such as silicon. On one side of the plane, the rod is doped with donor atoms (thus creating *n*-type material), and on the other side it is doped with acceptor atoms (thus creating *p*-type material). This combination gives a *pn* junction.

Figure 49-10*a* represents a *pn* junction at the imagined moment of its creation. There is an abundance of electrons in the *n*-type material and of holes in the *p*-type material.

Electrons close to the junction plane will tend to diffuse across it, for much the same reason that gas molecules will diffuse through a permeable membrane into a vacuum beyond it. The diffusing electrons in the conduction band, which move from right to left in Fig. 49-10*a*, will readily combine with the holes in the valence band on the other side of the junction plane. Similarly, the holes in the *p*-type region diffuse across the junction plane from left to right and combine with electrons in the *n*-region.

For every such diffusion–recombination event, the portion of the bar on the right side of this plane acquires a positive charge and the portion on the left side a negative charge. These charges* cause a potential difference ΔV_0 to build up across the junction, as Fig. 49-10*c* shows. Related to the potential difference (by the equation $E = -dV/dx$) is an internal electric field E_0 that appears across the junction plane, pointing as shown in Fig. 49-10*c*. This field exerts a force on the electrons, opposing their motion of diffusion. Put another way, for an electron to succeed in diffusing from right to left or a hole from left to right in Fig. 49-10*b*, it must be energetic enough to overcome the potential barrier ΔV_0 shown in Fig. 49-10*c*. This is represented in Fig. 49-10*d*, which shows the electron energy bands. To diffuse from the *n*-type region to the *p*-type region, an electron must "climb" the hill of height $e\Delta V_0$. A hole must also "climb" a hill of this same height to diffuse from left to right. The diffusion of both electrons and holes gives a current whose direction, in the usual conventional sense, is from left to right in Fig. 49-10. We call this current the *diffusion current* i_{diff}.

It is, of course, not possible to have an isolated silicon rod resting on a shelf with a current flowing indefinitely

---

*The fixed charges, which are close to—and separated by—the junction plane, are those of the donor and the acceptor ion cores, which, we recall, are not mobile. Normally the charges of these ion cores are compensated by the (opposite) charges of the mobile charge carriers. However, when charge carriers cross the junction plane, the ion cores are no longer fully compensated and are, so to speak, uncovered.

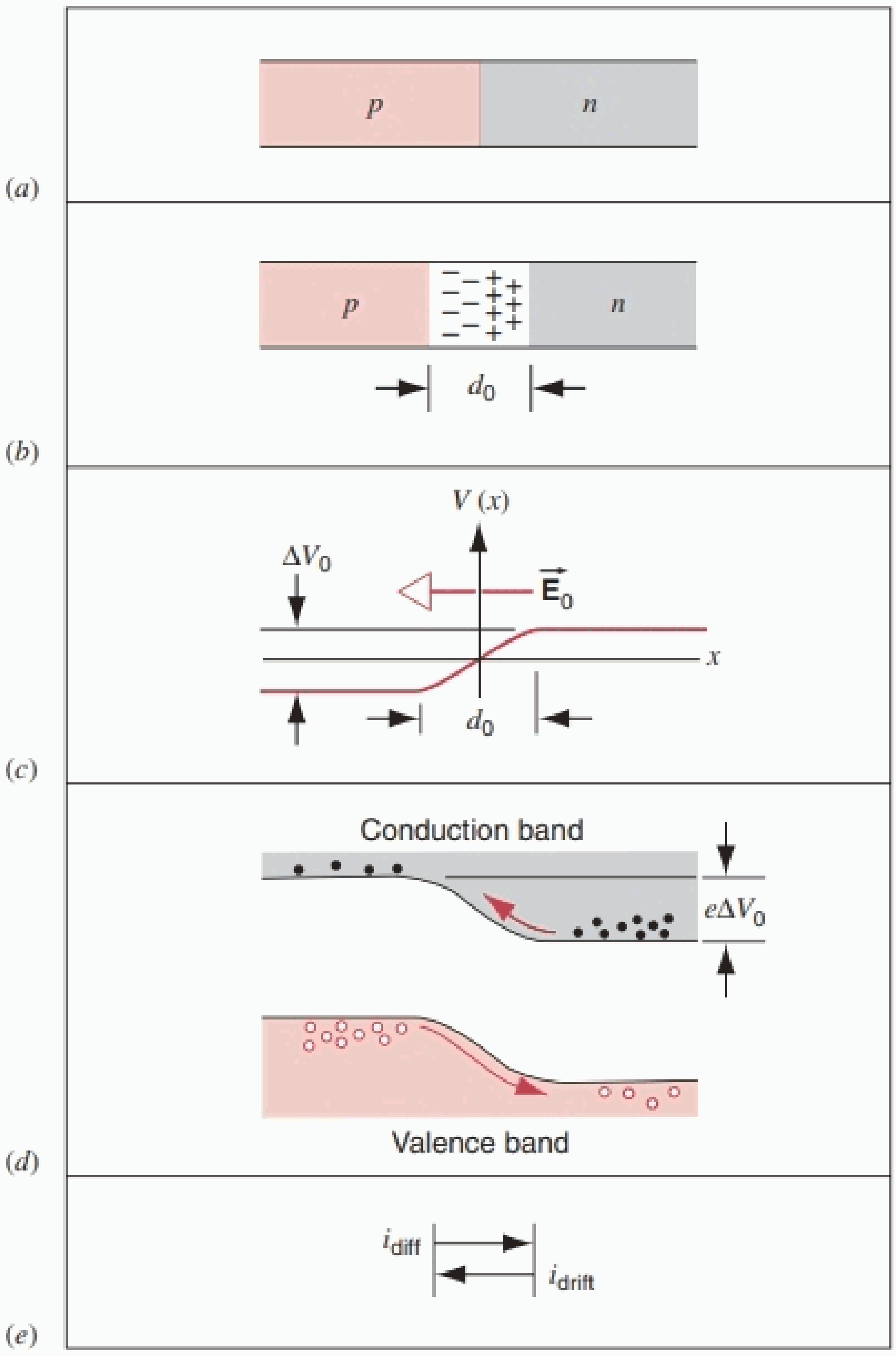


FIGURE 49-10. (*a*) A *pn* junction at the imagined moment of its creation. (*b*) Diffusion of majority carriers across the junction plane causes a space charge of fixed donor and acceptor ions to appear. (*c*) The space charge establishes a potential difference ΔV_0 and a corresponding electric field $\vec{\mathbf{E}}_0$ across the junction plane. (*d*) The electron energy bands near the junction. The arrows show the diffusion of the majority carriers. (*e*) In equilibrium, the diffusion of majority carriers across the junction plane is just balanced by the drift of minority carriers in the opposite direction.

along its length. Something must happen to stop, or to compensate, this current. To find out what it is, we turn our attention to the minority carriers.

As Fig. 49-9 and Table 49-2 show, although the majority carriers in *n*-type material are electrons, there are nevertheless also a few holes—the minority carriers. Likewise in *p*-type material, although the majority carriers are holes, there are also a few electrons. The minority carriers are shown in Fig. 49-10*d*.

Although the electric field in Fig. 49-10*c* acts to retard the motions of the majority carriers—being a barrier for them—it is a downhill trip for the minority carriers, be they electrons or holes. When, by thermal agitation, an electron close to the junction plane is raised from the valence band to the conduction band of the *p*-type material, it drifts steadily from left to right across the junction plane, swept along by the electric field E_0. Similarly, if a hole is created in the *n*-type material, it too drifts across to the other side. The space-charge region shown in Fig. 49-10*b* is effectively swept free of charge carriers by this process and, for that reason, we call it the *depletion zone.* The current represented by the motions of the minority carriers, called the *drift current* i_{drift}, is in the opposite direction to the diffusion current and exactly compensates it at equilibrium, as Fig. 49-10*e* shows.

Thus, at equilibrium, a *pn* junction resting on a shelf develops a contact potential difference ΔV_0 between its ends. The diffusion current i_{diff} that moves through the junction plane from the direction *p* to *n* is balanced by a drift current i_{drift} that moves in the opposite direction. An electric field E_0 acts across the depletion layer, whose width is d_0.

SAMPLE PROBLEM 49-9. A silicon-based *pn* junction has an equal concentration n_0 of donor and acceptor atoms. Its depletion zone, of width *d*, is symmetrical about the junction plane, as Fig. 49-11*a* shows. (*a*) Derive an expression for $E_{\max}$, the maximum intensity of the electric field in the depletion zone. (*b*) Derive an expression for ΔV_0, the potential difference that exists across the depletion zone; see Fig. 49-10*c*. (*c*) Assume that $n_o = 3 \times 10^{22}\ \text{m}^{-3}$ and that ΔV_0 is measured to be 0.6 V. Calculate the width of the depletion zone. (*d*) Using this value of *d*, calculate the value of $E_{\max}$.

Solution (*a*) The electric field may be taken as zero in the *n*-type and the *p*-type material outside the depletion zone. The field points from right to left within the depletion zone and, from symmetry, has its maximum value in the center of this zone; see Fig. 49-11*b*.

Let us apply Gauss' law to the closed "box" (Gaussian surface) shown in Fig. 49-11*a*. This law is

$$\epsilon_0 \oint \kappa_e \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = q,$$

in which κ_e (= 12) is the dielectric constant of silicon and q [$= n_0 eA(d/2)$] is the free charge contained within the box. The integral is to be taken over the surface of the box.

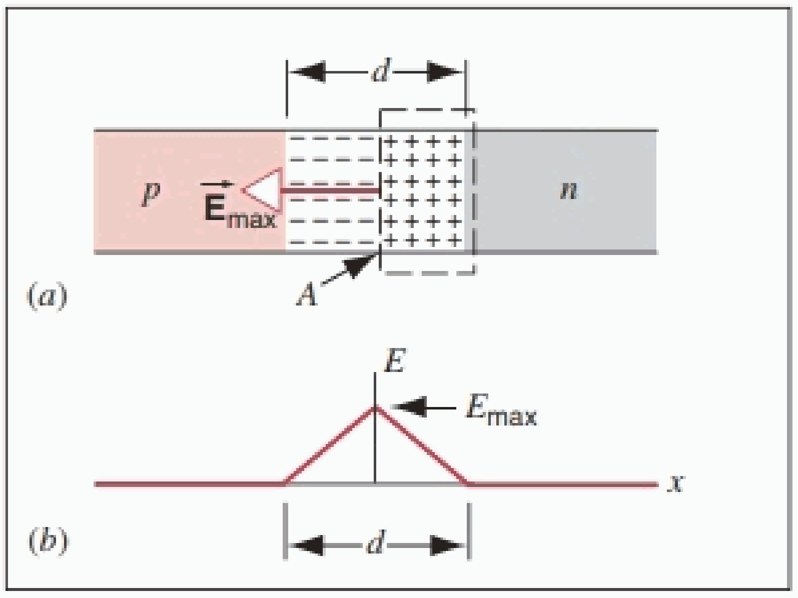


FIGURE 49-11. Sample Problem 49-9. (*a*) The depletion zone of a *pn* junction. The dashed rectangle represents the cross section of a Gaussian surface with end caps of area *A*. (*b*) The variation of the electric field in the depletion zone.

The only contribution to the integral comes from the face of the box that lies in the junction plane so that the integral has the value $\kappa_e E_{\max} A$. Making these substitutions and solving for $E_{\max}$ yields

$$E_{\max} = \frac{n_0 e d}{2\kappa_e \epsilon_0}, \qquad (49\text{-}9)$$

the relationship we seek.

(*b*) As Fig. 49-11*b* shows, the electric field drops linearly from its central value of $E_{\max}$ to zero at each edge of the depletion zone. Its *average* value throughout the zone is thus $\frac{1}{2}E_{\max}$. The potential difference ΔV_0 is equal to the work per unit charge required to carry a test charge q_0 from one face of the depletion zone to the other. Thus if F is the average force acting on the test charge,

$$\Delta V_0 = \frac{W}{q_0} = \frac{Fd}{q_0} = \frac{(\frac{1}{2}E_{\max} q_0)d}{q_0} = \tfrac{1}{2}E_{\max} d.$$

Substituting for $E_{\max}$ from Eq. 49-9 leads to

$$\Delta V_0 = \frac{n_0 e d^2}{4\kappa_e \epsilon_0}. \qquad (49\text{-}10)$$

(*c*) Solving Eq. 49-10 for d and substituting the given values, we find

$$\begin{aligned} d &= \left(\frac{4\kappa_e \epsilon_0 \,\Delta V_0}{e n_0}\right)^{1/2} \\ &= \left[\frac{(4)(12)(8.85 \times 10^{-12}\text{ F/m})(0.60\text{ V})}{(1.60 \times 10^{-19}\text{ C})(3 \times 10^{22}\text{ m}^{-3})}\right]^{1/2} \\ &= 2.3 \times 10^{-7}\text{ m} = 230\text{ nm}. \end{aligned}$$

(*d*) Substituting in Eq. 49-9 leads to

$$\begin{aligned} E_{\max} &= \frac{n_0 e d}{2\kappa_e \epsilon_0} \\ &= \frac{(3 \times 10^{22}\text{ m}^{-3})(1.60 \times 10^{-19}\text{ C})(2.3 \times 10^{-7}\text{ m})}{(2)(12)(8.85 \times 10^{-12}\text{ F/m})} \\ &= 5.2 \times 10^{6}\text{ V/m}. \end{aligned}$$

What assumptions were made in this problem that might lead to different values of the calculated quantities under practical laboratory conditions?

The Diode Rectifier

Although a *pn* junction can be used in many ways, it is basically a rectifier. That is, if you connect it across the terminals of a battery, the current (a few picoamperes) in the circuit will be very much smaller for one polarity of the battery connection than for the other. Figure 49-12 shows that, for a typical silicon-based *pn* junction diode, the current for the *reverse-biased* connection ($\Delta V < 0$) is negligible by comparison with the current for the forward-biased connection ($\Delta V > 0$).

Figure 49-13 shows one of many possible applications of a diode rectifier. A sine wave input potential generates a half-wave output potential, the diode rectifier acting as essentially a short circuit for one polarity of the input potential and as essentially an open circuit for the other. An ideal diode rectifier, in fact, has only these two modes of operation. It is either on (zero resistance) or off (infinite resistance).

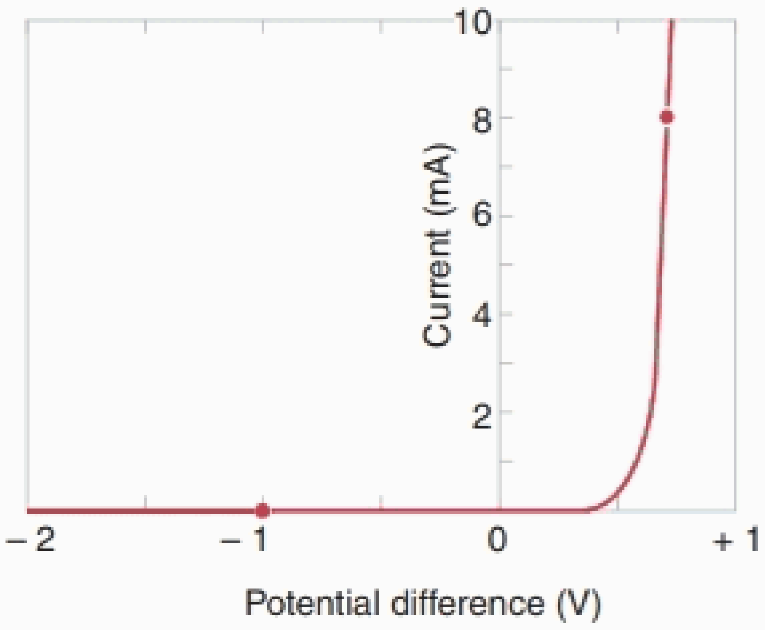


FIGURE 49-12. A current–voltage plot for a typical *pn* junction, showing that it conducts easily in the forward direction but is essentially nonconducting in the reverse direction. The dots refer to Exercise 35.

Figure 49-13 displays the conventional symbol for a diode rectifier. The arrow head corresponds to the *p*-type terminal of the device and points in the direction of "easy" conventional current flow. That is, the diode is on when the terminal with the arrow head is (sufficiently) positive with respect to the other terminal.

Figure 49-14 shows details of the two connections. In Fig. 49-14*a* (the reverse-biased arrangement) the battery emf simply adds to the contact potential difference, thus increasing the height of the barrier that the majority carriers must surmount. Fewer of them can do so and, as a result, the diffusion current decreases markedly.

The drift current, however, senses no barrier and thus is independent of the magnitude or direction of the external potential. The current balance that existed at zero bias (see Fig. 49-10*e*) is thus upset and, as shown in Fig. 49-14*a*, a current—but a very small one—appears in the circuit.

Another effect of reverse bias is to widen the depletion zone, as a comparison of Figs. 49-10*b* and 49-14*a* shows. This seems reasonable because the positive battery terminal, connected to the *n*-type end of the junction, tends to pull electrons out of the depletion zone back into the *n*-type material and to repel holes back into the *p*-type material. Because the depletion zone contains very few charge carriers, it is a region of high resistivity. Thus its substantially increased width means a substantially increased resistance, consistent with the small value of the current in a reverse-biased diode.

Figure 49-14*b* shows the forward-biased connection, with the positive terminal of the battery connected to the *p*-type end of the *pn* junction. Here the applied emf *subtracts* from the contact potential, the diffusion current *rises* substantially, and a relatively *large* net forward current results. The depletion zone becomes *narrower,* its low resistance consistent with the large forward current.

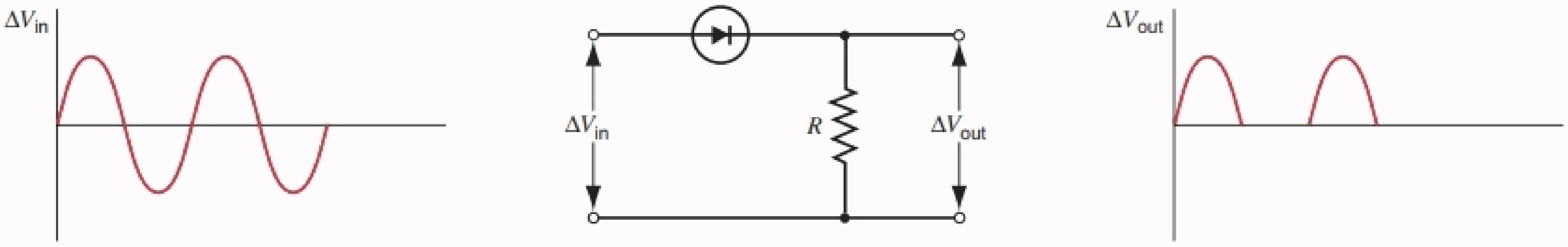


FIGURE 49-13. A *pn* junction diode connected as a rectifier. The diode conducts easily in the forward direction (positive sections of input wave) but not at all in the reverse direction (negative sections of input wave).

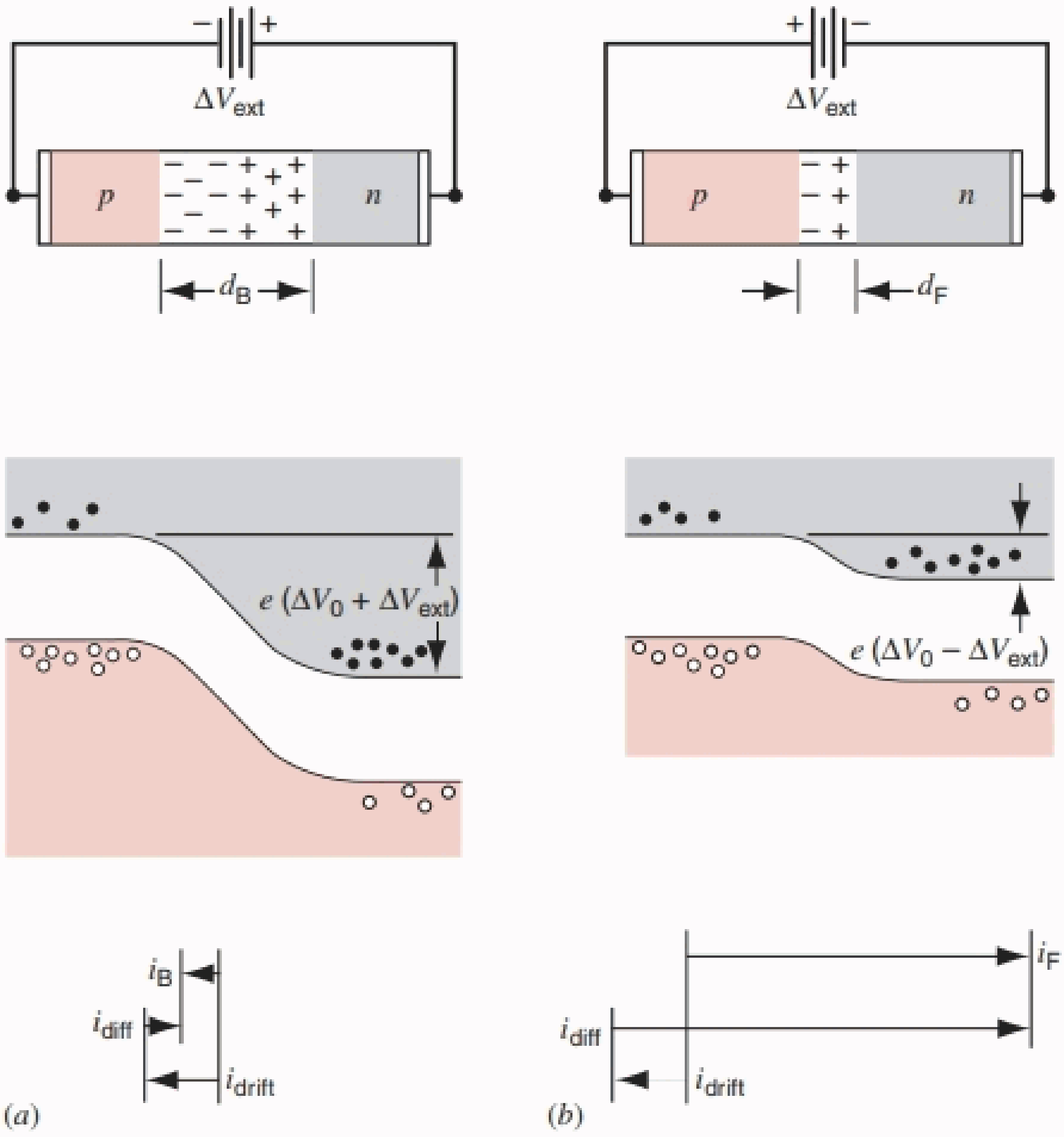


FIGURE 49-14. (*a*) The reverse-biased connection of a *pn* junction, showing the wide depletion zone, the energy bands, and the corresponding small back current i_B. (*b*) The forward-biased connection, showing the narrow depletion zone, the energy bands, and the large forward current i_F. Note that the drift current is the same in each case.

49-9 OPTICAL ELECTRONICS

We are all familiar with the brightly colored numbers that flash and glow at us from cash registers, gasoline pumps, and electronic equipment. In nearly all cases this light is emitted from an assembly of *pn* junctions operating as *light-emitting diodes* (LEDs).

Figure 49-15*a* shows the familiar seven-segment display from which the numbers are formed. Figure 49-15*b* shows that each element of this display is the end of a flat plastic lens, at the other end of which is a small LED, possibly about 1 mm^2 in area. Figure 49-15*c* shows a typical circuit, in which the LED is forward biased.

How can a *pn* junction emit light? When an electron at the bottom of the conduction band of a semiconductor falls into a hole at the top of the valence band, an energy E_g is released, where E_g is the gap width. What happens to this energy? There are at least two possibilities. It might be transformed into internal energy of the vibrating lattice and, with high probability, that is exactly what happens in a silicon-based semiconductor.

In some semiconducting materials, however, the emitted energy can appear as electromagnetic radiation, with the wavelength given by

$$\lambda = \frac{c}{f} = \frac{c}{E_g/h} = \frac{hc}{E_g}. \tag{49-11}$$

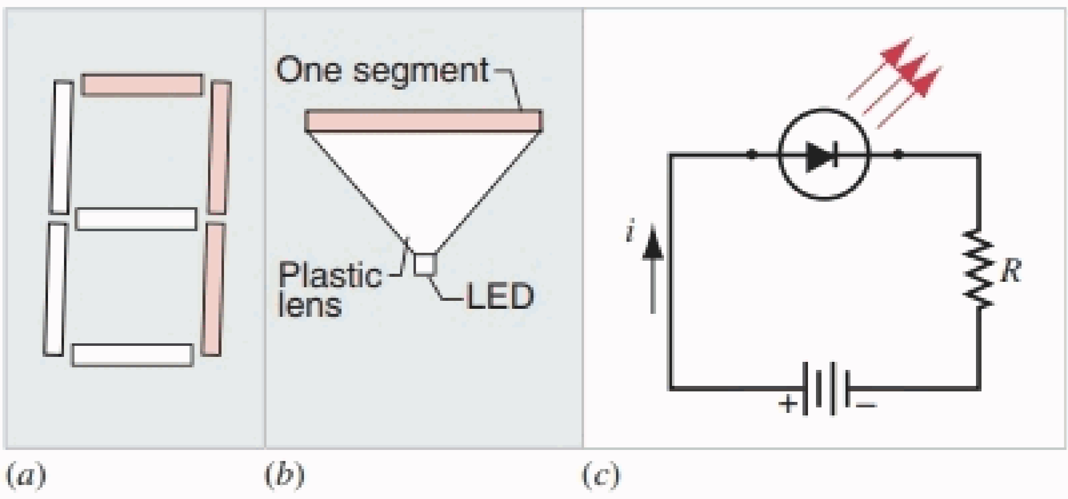


FIGURE 49-15. (*a*) The familiar seven-segment number display, activated to show the number "7." (*b*) One segment of such a display. (*c*) An LED connected to an external source of emf.

Commercial LEDs designed for the visible region are usually based on a semiconducting material that is a gallium–arsenic–phosphorus compound. By adjusting the ratio of phosphorus to arsenic, the gap width—and thus the wavelength of the emitted light—can be varied.

If light is emitted when an electron falls from the conduction band to the valence band, then light of that same wavelength will be *absorbed* when an electron moves in the other direction—that is, from the valence band to the conduction band. To avoid having all the emitted photons absorbed, it is necessary to have a great surplus of both electrons and holes present in the material, in much greater numbers than would be generated by thermal agitation in the intrinsic semiconducting material. These are precisely the conditions that result when majority carriers—be they electrons or holes—are injected across the central plane of a *pn* junction by the action of an external potential difference. That is why a simple intrinsic semiconductor will not serve as an LED. You need a *pn* junction! To provide lots of majority carriers—and thus lots of photons—it should be heavily doped and strongly forward biased.

SAMPLE PROBLEM 49-10. An LED is constructed from a *pn* junction based on a certain semiconducting material whose energy gap is 1.90 eV. What is the wavelength of its emitted light?

Solution From Eq. 49-11 we have

$$\lambda = \frac{hc}{E_g} = \frac{(6.63 \times 10^{-34}\,\mathrm{J\cdot s})(3.00 \times 10^8\,\mathrm{m/s})}{(1.90\,\mathrm{eV})(1.60 \times 10^{-19}\,\mathrm{J/eV})}$$

$$= 6.54 \times 10^{-7}\,\mathrm{m} = 654\,\mathrm{nm}.$$

Light of this wavelength is red.

The Diode Laser

The dropping down of an electron from the conduction band to fill a hole in the valence band, with the emission of a photon, bears a strong resemblance to the dropping down of electrons in transitions between atomic states we considered in Chapter 48. There is an important application based on this similarity: by injecting electrons into the conduction band and holes into the valence band, it is possible to create a population inversion analogous to that considered in our discussion of lasers in Section 48-9. In this way it is possible to make a *diode laser,* in which the lasing medium is not a gas but a solid semiconductor. Diode lasers are commonly used in compact disk players and other optical data retrieval systems.

Figure 49-16 shows a representation of the energy levels in a diode laser. The lasing material is sandwiched between layers of *p*-type and *n*-type material, which have slightly larger gap energies. Electrons are injected by an external circuit into the *n*-type material; some of these excess electrons drift into the lasing layer, where they are prevented from drifting into the *p*-type material by a potential barrier. Similarly, holes are injected into the *p*-type material, drift into the lasing layer, and are trapped there. The excess of electrons (and holes) in the active region gives the lasing action.

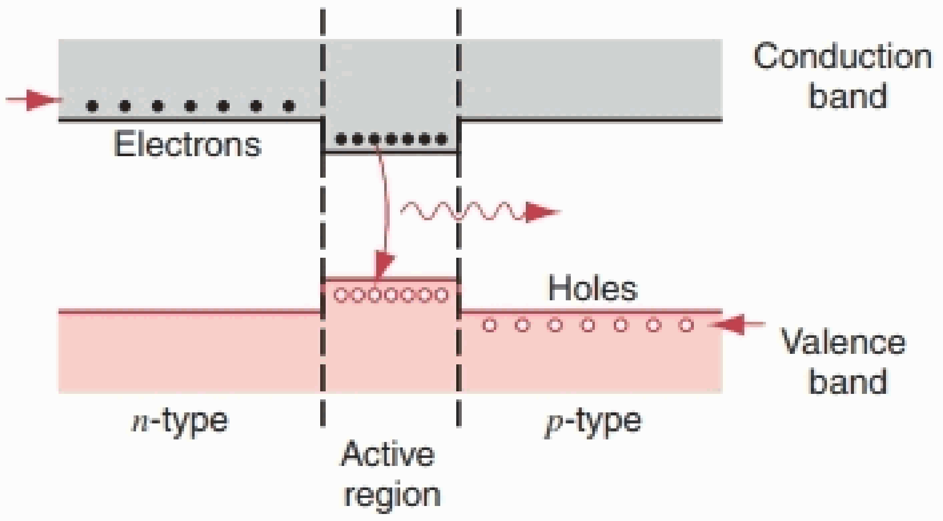


FIGURE 49-16. The energy bands in a diode laser. The active region has a smaller energy gap than the *n*-type and *p*-type materials on either side. When electrons in the conduction band of the active region drop down to fill holes in the valence band, light is emitted.

The physical construction of the device is illustrated schematically in Fig. 49-17, and Fig. 49-18 shows a photograph of a diode laser. The lasing material is a narrow (0.2 μm) layer of a material such as GaAs (gallium arsenide), and the *p*-type and *n*-type material on each side may be layers of GaAlAs (gallium aluminum arsenide) a few micrometers in thickness. The ends of the material are cleaved to create mirror-like surfaces that reflect a portion of the light wave to enable stimulated emission in the active region. Diode lasers that emit at infrared wavelengths are commonly used in communication to send signals along optical fibers. Other materials can be used in similar fashion to give visible radiation.

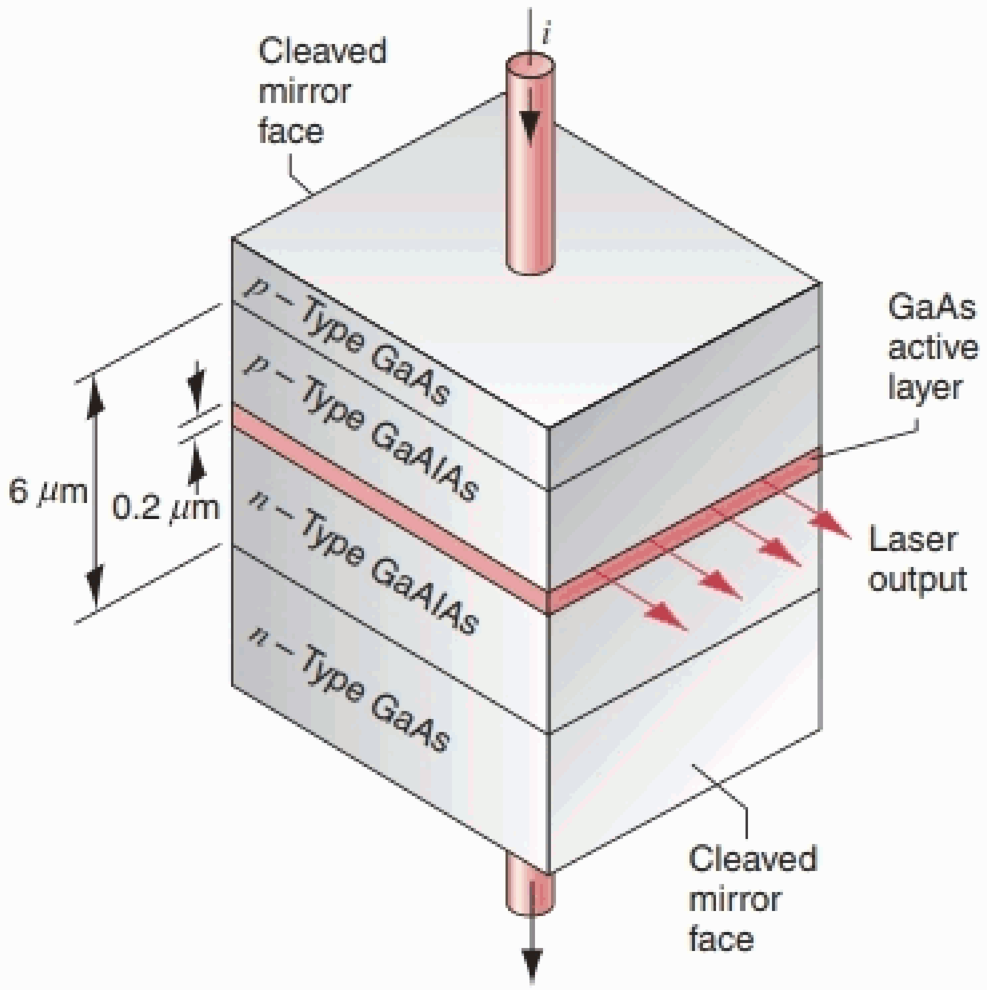


FIGURE 49-17. The physical construction of a diode laser. The lasing action occurs in the narrow GaAs layer.

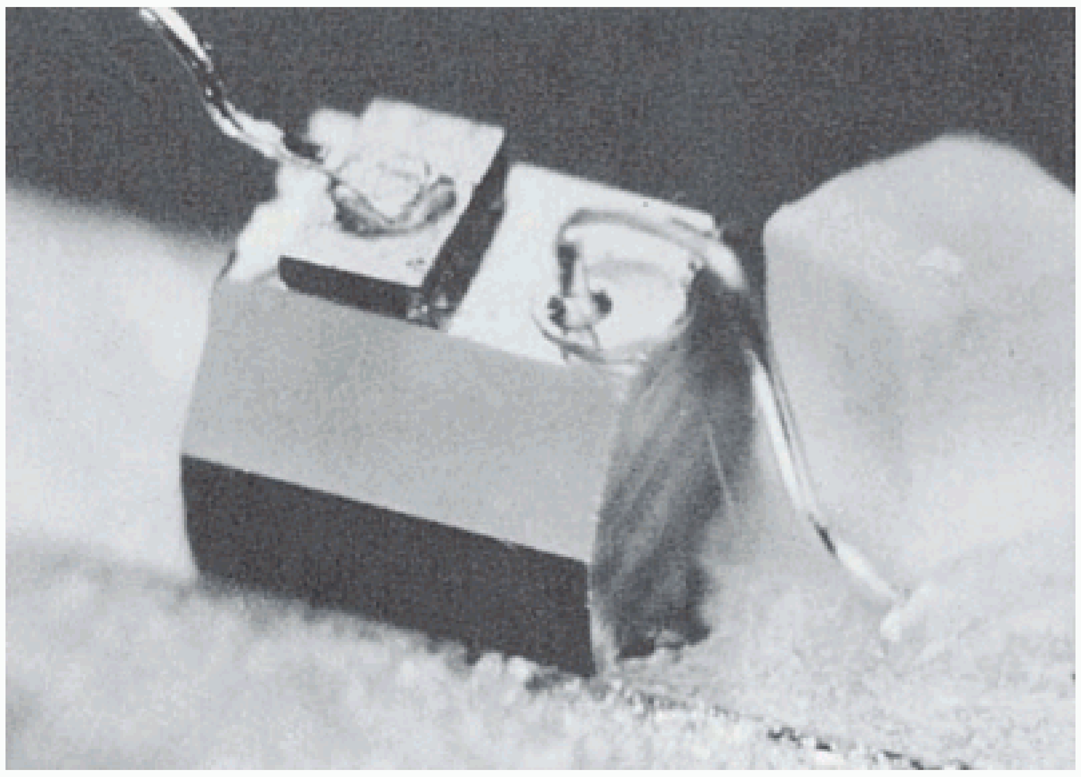


FIGURE 49-18. A diode laser, compared in size with a grain of table salt on the right.

Among the advantages of diode lasers are their small size and low power input (in the range of 10 milliwatts, compared with the standard HeNe laser that may require several watts of electrical power). Like other semiconductor devices, the diode laser can be powered by batteries. Efficiencies of the order of 20% are possible (that is, 20% of the electrical power supplied to the device appears in the laser beam), compared with 0.1% in the HeNe laser. The light signal can easily be modulated by controlling the injection current, and thus we have an optical device that can respond at the rapid switching times (< 100 ps) characteristic of electronic circuits.

49-10 THE TRANSISTOR

The junction diodes we have considered so far are two-terminal devices. Here we consider a device with three (or more) terminals, called a *transistor.** A transistor often operates in the following mode: a current established between two of the terminals is regulated by a current or voltage at the third terminal.

One common variety of transistor is the *junction transistor,* which consists of three layers of doped semiconductors, such as *npn* or *pnp.* Figure 49-19*a* shows a typical configuration for an *npn* junction transistor. The three sections are called the *emitter, base,* and *collector.* The conduction and valence bands are shown, and only the majority carriers are indicated. The emitter–base and base–collector junctions behave much like ordinary *pn* junctions.

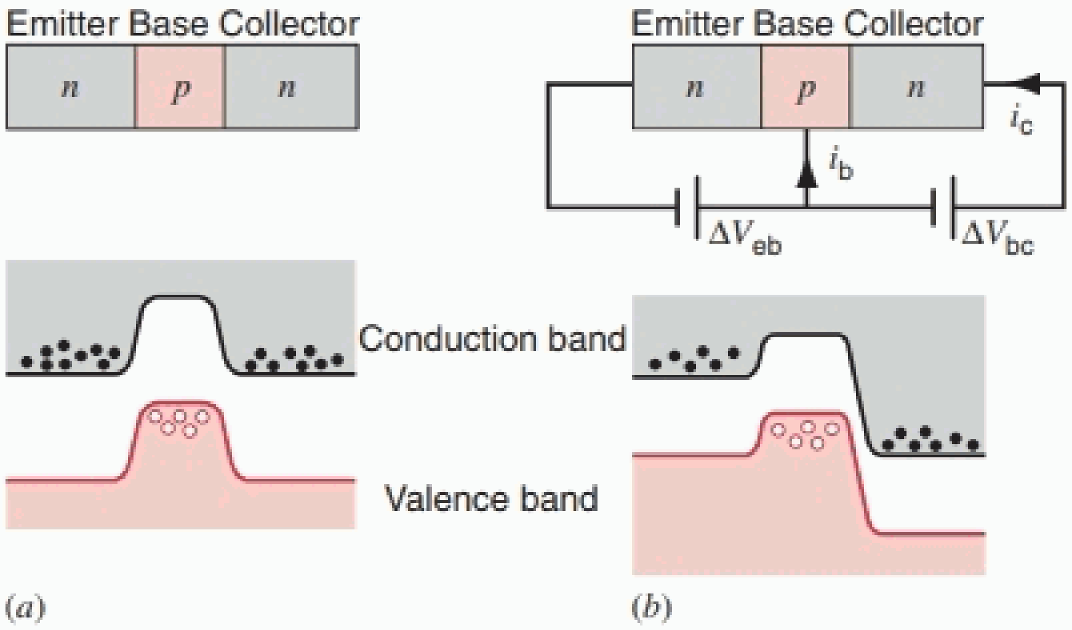


FIGURE 49-19. (*a*) An *npn* junction transistor. At the bottom are shown the energy bands and majority carriers in the three regions. (*b*) The emitter–base junction is forward biased, and the base–collector junction is reverse biased. Electrons that move from the emitter to the base either recombine with holes or (far more likely) continue to the collector.

In normal operation, as illustrated in Fig. 49-19*b*, the emitter–base junction is forward biased and the base–collector junction is reverse biased, which gives the energy bands shown in the figure.

Electrons flow from the heavily doped *n*-type emitter into the base. Because the base is very narrow, most of these electrons reach the collector, but a few recombine with holes in the *p*-type region. To replenish the holes in the base, electrons from the valence band in the base must leave the transistor through the external circuit as the small base current i_b. A small change in the base current i_b can result in a large change in the collector current i_c. In this configuration, the transistor serves as a current amplifier, and the current gain i_c/i_b can have typical values in excess of 100.

A second type of transistor is called a *field-effect transistor* (FET). Figure 49-20 illustrates the basic geometry. Electrons flow through the *n*-type region from the *source* to the *drain* when there is an external potential difference ΔV_{ds} between the drain and the source. The *p*-type regions are heavily doped, and the depletion layers formed at the two *pn* junctions determine the width of the *n*-type channel. An external voltage ΔV_g applied to the *p*-type region (the *gate*) changes the width of the depletion region and consequently changes the width of the *n*-type channel. This in

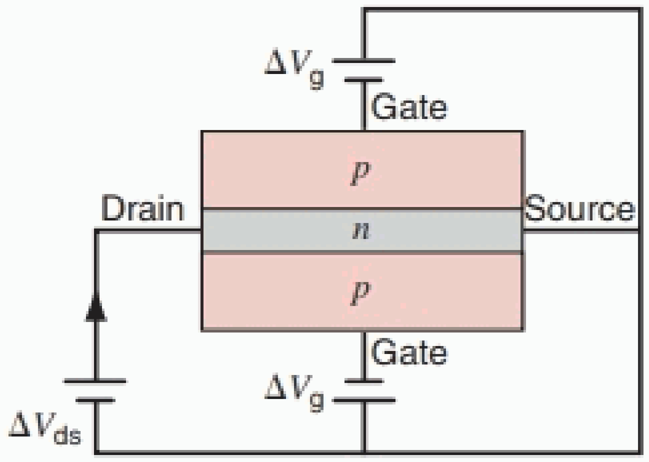


FIGURE 49-20. The basic structure of a field-effect transistor. Electrons travel along the narrow *n*-channel from the source to the drain. The width of the channel can be controlled by varying the voltage ΔV_g at the gate.

---

*The transistor was invented in 1947 at what is now the Lucent Technologies (formerly AT&T) Bell Laboratories by John Bardeen, Walter Brattain, and William Shockley, who shared the 1956 Nobel Prize in physics for their discovery.

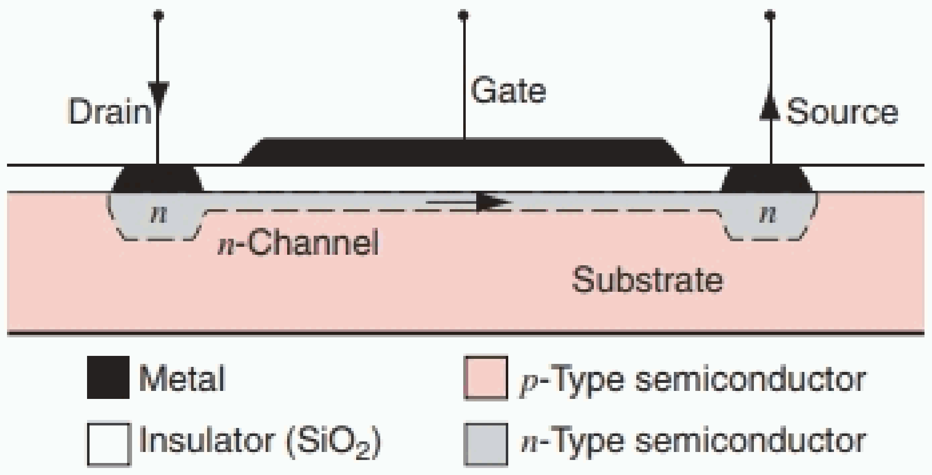


FIGURE 49-21. The structure of a MOSFET.

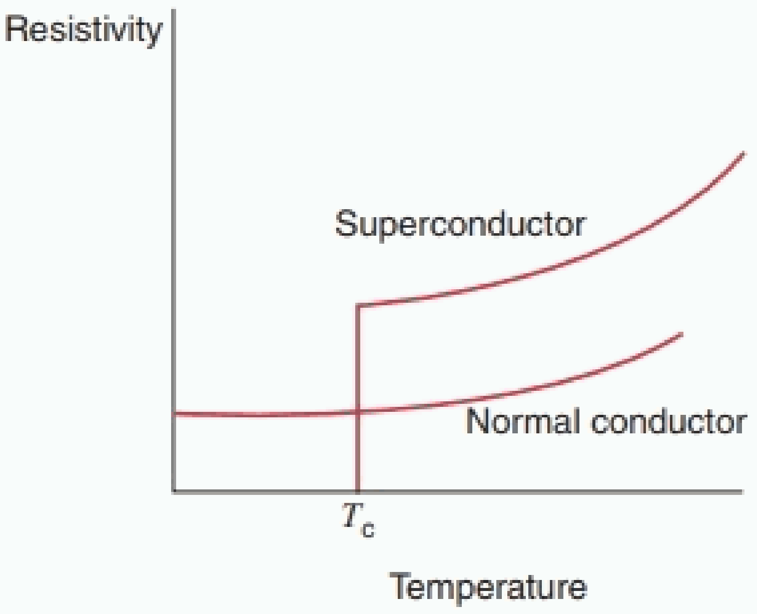


FIGURE 49-22. Comparison of the dependence of resistivity on temperature for a normal conductor and a superconductor. The resistivity of a normal conductor falls gradually with decreasing temperature. In superconducting materials, the resistivity drops suddenly to zero at the critical temperature T_c.

turn changes the current through the device, because the ability of current to flow along the *n*-channel depends on the width of the channel. A small change in the gate voltage changes the width of the channel and causes a large change in the current through the *n*-channel so the device can operate as an amplifier.

If the gate voltage is made large enough, the *n*-channel width can become zero, and the FET stops conducting. Here the transistor is acting like a switch: it is either conducting (on) or not conducting (off). The current can be switched on or off very rapidly by the signal applied to the gate; switching times smaller than 1 ns (10^{-9} s) are common.

A common type of FET widely used in digital circuits is the metal-oxide-semiconductor FET (MOSFET), which is fabricated by depositing and etching successive layers in a *p*-type substrate. A cross section of a MOSFET is shown in Fig. 49-21. The *n*-region and the *n*-channel are made by etching a mask onto the *p*-type substrate and diffusing donor atoms a known distance into the substrate. An oxide layer (SiO_2) is then deposited, and a metal layer is then deposited to form the contacts for the *n*-region and the gate.

49-11 SUPERCONDUCTORS

As we reduce the temperature of a conductor, the resistivity grows smaller, as Fig. 29-8 suggests. What happens as we approach the absolute zero of the temperature scale?

The part of the resistivity due to scattering of electrons by atoms vibrating from their equilibrium lattice positions decreases as the temperature decreases, because the amplitude of the vibration decreases with temperature. According to quantum theory, the atoms retain a certain minimum vibrational motion, even at the absolute zero of temperature. Furthermore, the contributions of defects and impurities to the resistivity remain as T falls to 0. We therefore expect the resistivity to decrease with decreasing temperature, but to remain finite at the lowest temperatures. Many materials do in fact show this type of behavior, as illustrated in Fig. 49-22.

Quite a different kind of behavior was discovered in 1911 by the Dutch physicist Kammerlingh Onnes, who was studying the resistivity of mercury at low temperature. He discovered that, below a temperature of about 4 K, mercury suddenly lost all resistivity and became a *perfect* conductor, called a *superconductor.* This was not a gradual change, but a sudden transition, as indicated by Fig. 49-22. The resistivity of a superconductor is not only very small; it is zero! If a current is established in a superconducting material, it should persist forever, even with no electric field present. The temperature at which a material becomes superconducting is called its *critical temperature* T_c. Table 49-3 shows some superconducting materials and their critical temperatures.

Superconductivity has been observed in 27 elements and in numerous compounds, but it has not been observed for the best metallic conductors (Cu, Ag, Au). We conclude that a superconductor is not merely a good conductor getting better, and we are led to suspect that the mechanism that causes superconductivity may differ from the mechanism that causes the conductivity of ordinary metals. As we shall see, superconductivity results from a *strong* coupling between conduction electrons and the lattice. Normal conduction in the best conductors occurs when there is a *weak* coupling between the valence electron and the lattice.

TABLE 49-3 **Properties of Some Superconductors**

Material	T_c (K)	*Pairing Energy* (meV)
Cd	0.56	0.27
Al	1.19	0.34
Sn	3.75	1.15
Hg	4.16	1.65
Pb	7.22	2.73
Nb	9.46	3.05
Nb_3Sn	18.1	
MgB_2	39	
$YBa_2Cu_3O_7$	90	
$Hg_{.8}Tl_{.2}Ba_2Ca_2Cu_3O_{8.33}$	138	

Consider an electron moving through a lattice. As it moves, it pulls the positive ion cores toward it and changes the charge density in its vicinity. It leaves a somewhat higher positive charge density in its wake than would otherwise be there. This positive charge attracts other electrons. The electrons interact with one another through the intermediary of the lattice, somewhat like two boats on a lake interacting through their wakes. The net result is a slight attraction of the electrons for each other.

The BCS (Bardeen–Cooper–Schrieffer) theory* of superconductivity shows that the electron system has the lowest possible energy if the electrons are bound together in pairs, called *Cooper pairs.* When no current exists in a superconductor, the two electrons of a Cooper pair have momenta of equal magnitude but exactly opposite directions, so that the total momentum and the electric current both vanish. When a current is generated, both electrons in a pair acquire the same increase in momentum, resulting in a motion of the center of mass of the pair. All Cooper pairs acquire the same momentum.

Superconductivity is a cooperative phenomenon. If some Cooper pairs have been formed, the reduction in energy that occurs for the next pair is greater than if no pairs were previously formed. Once the temperature drops below T_c and some pairs are formed, a small additional reduction in temperature causes many more pairs to form. The change from the normal to the superconducting state is quite precipitous. The cooperative motions of the Cooper pairs also force every pair to have the same momentum.

The Cooper pairs have a binding energy Δ, called the *pairing energy,* which is typically in the range of 10^{-4} to 10^{-3} eV, as shown in Table 49-3. Note that critical temperatures of $1-10$ K (typical for most of the superconductors shown in Table 49-3) correspond to energies kT_c in the same range of 10^{-4} to 10^{-3} eV. The critical temperature of a superconductor is directly related to the pairing energy. Above T_c, the pairs are broken and the material has normal electrical resistance.

The binding energy of a Cooper pair introduces a *pairing gap* 2Δ into the density of states $n(E)$ near the Fermi energy. (Figure 49-1*a* shows an example of the density of states for a normal conductor.) It is energetically favorable for electrons near the Fermi energy in a superconductor to bind together in Cooper pairs. As a result, the density of states decreases to zero within an interval of $\pm\Delta$ of E_F, with a corresponding increase in $n(E)$ just above and below E_F. Figure 49-23 shows the resulting density of states and the pairing gap 2Δ. Above T_c, the density of states of a superconductor might be as shown in Fig. 49-2*a*. The gap begins to open as the superconductor is cooled below T_c; the gap energy increases as the temperature decreases, reaching its maximum as T approaches 0 K.

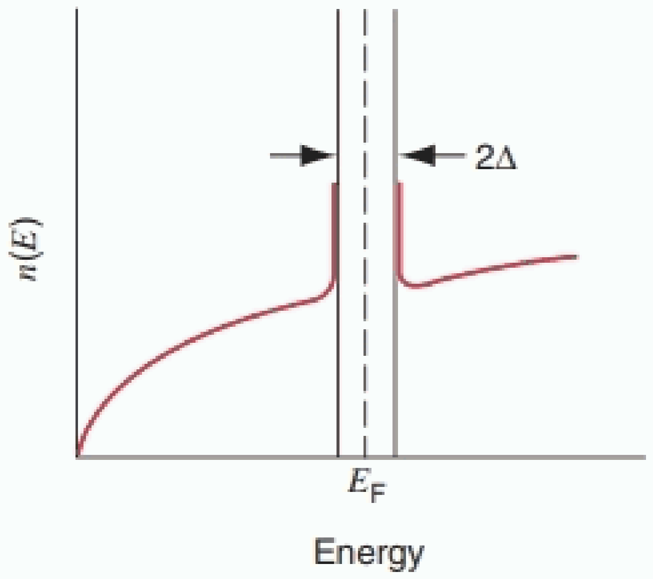


FIGURE 49-23. The density of states in a superconductor below its transition temperature. There is an energy gap of 2Δ, within which the density of states is zero. The scale of this drawing has been exaggerated; typically the Fermi energy E_F is a few electron-volts, while the pairing gap is 10^{-4} to 10^{-3} eV.

The occupation probability of electron states in a superconductor can be found from the product of the density of states, shown in Fig. 49-23, and a Fermi–Dirac distribution function, as was shown in Fig. 49-2*b*. This leads to a high occupation probability of the superconducting states just below the gap. Above the gap, a small density of normal (unpaired) states occurs for $T > 0$.

Beginning in 1986, a new class of superconductors was discovered* with unusually high values of T_c. The last two entries in Table 49-3 are examples of these compounds, which are ceramic materials that (unlike the more familiar types of ceramics) are conductors at room temperature. Since the highest temperature at which superconductivity had previously been observed was about 20 K, these new materials represent a substantial leap in technology. In particular, they allow superconductivity to be attained at temperatures characteristic of cooling with liquid nitrogen (77 K) rather than at those characteristic of the more expensive and less convenient liquid helium (4 K). This jump of a factor of 6 in T_c holds out the hope that, with another jump of a factor of less than 3, it might be possible to achieve superconductivity at room temperature.

These high-temperature superconductors are oxides of copper in combination with various other elements. The theory of operation of these materials is not yet understood; it is not clear whether there is a BCS-type mechanism involved. It seems apparent that the superconductivity resides with the copper oxides; although elemental copper is not superconducting, the copper oxide combinations are. The crystal structure of these compounds places the copper and oxygen in planes anchored between the other elements, and it is likely that these planes provide the pathway for the electrons that carry the superconducting current.

---

*This theory of superconductivity was developed in 1957 by John Bardeen, Leon N. Cooper, and J. Robert Schrieffer, who were awarded the 1972 Nobel Prize in physics for their work. Bardeen also shared the 1956 Nobel Prize for his research on semiconductors and the discovery of the transistor.

*See "Superconductors Beyond 1-2-3," by Robert J. Cava, *Scientific American,* August 1990, p. 42.

The availability of superconducting materials immediately suggests a number of applications. (1) Energy can be transported and stored in electrical wires without resistive losses. That is, a power company can produce electrical energy when the demand is light, perhaps overnight, and store the current in a superconducting ring. Electrical power can then be released during peak demand times the following day. In smaller laboratory test rings, currents have been stored for several years with no reduction. (2) Superconducting electromagnets can carry larger currents and thus produce larger magnetic fields than conventional electromagnets. Applications of this technology include magnetically levitated trains and bending magnets for beams of particles in large accelerators such as Fermilab. (3) Superconducting components in electronic circuits would generate no Joule heating and would permit further miniaturization of circuits. The next generation of mainframe computers may employ superconducting components.

MULTIPLE CHOICE

49-1 Quantum Theory of Solids

49-2 Conduction Electrons in a Metal

49-3 Filling the Allowed States

1. The Fermi energy of copper at 0 K represents
(A) the average energy of the electrons.
(B) the energy necessary to remove an electron from the surface.
(C) the maximum energy of any electron.
(D) the minimum energy of any electron.

2. In a metal at 0 K, the probability of a state below the Fermi level being occupied is
(A) always 0. (B) always 1.
(C) dependent on its energy. (D) always 0.5.

3. In copper at room temperature, the probability of a state 1 eV above the Fermi level being occupied is about
(A) 1. (B) 0.5. (C) 0.025. (D) 10^{-18}.

49-4 Electrical Conduction in Metals

4. The speed of an electron at the Fermi surface of a typical metal is closest to
(A) the speed of light.
(B) the speed of sound in air.
(C) the drift speed of an electron in the metal.

49-5 Bands and Gaps

5. The existence of energy bands is a consequence of
(A) the uncertainty principle.
(B) interference of electron waves.
(C) random motions of the electrons.
(D) the interactions of atoms.

49-6 Conductors, Insulators, and Semiconductors

6. Specify which of the following types of materials might satisfy these descriptions: (*a*) a filled valence band, an empty conduction band, an energy gap of 8 eV; (*b*) a filled valence band, an empty conduction band, an energy gap of 1 eV; (*c*) a half-filled valence band, an empty conduction band, an energy gap of 1 eV.
(A) Conductor (B) Insulator
(C) Semiconductor (D) None of the above

7. For which type of material is the Fermi energy level located in the gap between the valence and conduction bands?
(A) Conductor and semiconductor
(B) Insulator and semiconductor
(C) Conductor only (D) Semiconductor only

49-7 Doped Semiconductors

8. You are given two samples of silicon, one *p*-type and the other *n*-type. How can you tell them apart?
(A) Measure their electrical resistivity.
(B) Measure their temperature coefficient of resistivity.
(C) Determine which is attracted by an object with a net positive charge.
(D) Determine which is attracted by the north pole of a magnet.
(E) None of the above

9. To make ordinary silicon into a *p*-type semiconductor, you might add to it
(A) indium (B) antimony.
(C) germanium. (D) carbon.

49-8 The *pn* Junction

10. When a forward bias is applied to a *pn* junction, the drift current
(A) increases. (B) decreases to zero.
(C) decreases but not to zero. (D) remains unchanged.

11. You have three identical *pn* junctions: junction 1 is unbiased, junction 2 is reverse biased, and junction 3 is forward biased. From largest to smallest, rank the three junctions according to the magnitude of their (*a*) drift currents, (*b*) diffusion currents, (*c*) net currents.
(A) 1, 2, 3 (B) 3, 1, 2
(C) 3, 2, 1 (D) 2, 3, 1
(E) All are the same.

49-9 Optical Electronics

49-10 The Transistor

49-11 Superconductors

12. Why is it necessary to cool a current-carrying superconductor to keep its temperature below T_c?
(A) Electrons form Cooper pairs only below T_c.
(B) The large currents result in significant Joule heating.
(C) The substance becomes superconducting only when kT is less than E_F.
(D) Both A and B are correct.

QUESTIONS

1. Do you think that any of the properties of solids listed in the opening to this chapter are related to each other? If so, which?
2. The conduction electrons in a metallic sphere occupy states of quantized energy. Does the average energy interval between adjacent states depend on (*a*) the material of which the sphere is made, (*b*) the radius of the sphere, (*c*) the energy of the state, or (*d*) the temperature of the sphere?
3. What role does the Pauli exclusion principle play in accounting for the electrical conductivity of a metal?
4. In what ways do the classical model and the quantum mechanical model for the electrical conductivity of a metal differ?
5. If we compare the conduction electrons of a metal with the atoms of an ideal gas, we are surprised to note (see Fig. 49-1*c*) that so much kinetic energy is locked into the conduction electron system at absolute zero. Would it be better to compare the conduction electrons, not with the atoms of a gas, but with the inner electrons of a heavy atom? After all, a lot of kinetic energy is also locked up in this case, and we do not seem to find that surprising. Discuss.
6. What features of Fig. 49-2 make it specific for copper, for which it was drawn? What features are independent of the identity of the metal?
7. Why do the curves in Figs. 49-1*c* and 49-2*c* differ so little from each other?
8. Distinguish carefully among the density of states function $n(E)$, the density of occupied states function $n_o(E)$, and the Fermi–Dirac probability function $p(E)$, all of which appear in Eq. 49-3.
9. Does the Fermi energy for a given metal depend on the volume of the sample? If, for example, you compare a sample whose volume is 1 cm^3 with one whose volume is twice that, the latter sample has just twice as many available conduction electrons; it might seem that you would have to go to higher energies to fill its available levels. Do you?
10. In Section 23-7 we showed that the (molar) heat capacity of an ideal monatomic gas is $\frac{3}{2}R$. If the conduction electrons in a metal behaved like such a gas, we would expect them to make a contribution of about this amount to the measured specific heat of a metal. However, this measured specific heat can be accounted for quite well in terms of energy absorbed by the vibrations of the ion cores that form the metallic lattice. The electrons do not seem to absorb much energy as the temperature of the specimen is increased. How does Fig. 49-2 provide an explanation of this prequantum-days puzzle?
11. Give a physical argument to account qualitatively for the existence of allowed and forbidden energy bands in solids.
12. Is the existence of a forbidden energy gap in an insulator any harder to accept than the existence of forbidden energies for an electron in, say, the hydrogen atom?
13. On the band theory picture, what are the *essential* requirements for a solid to be (*a*) a metal, (*b*) an insulator, or (*c*) a semiconductor?
14. What can band theory tell us about solids that the classical model (see Section 29-5) cannot?
15. Distinguish between the drift speed and the Fermi speed of the conduction electrons in a metal.
16. Why is it that, in a solid, the allowed bands become wider as one proceeds from the inner to the outer atomic electrons?
17. Do pure (undoped) semiconductors obey Ohm's law?
18. At room temperature a given applied electric field will generate a drift speed for the conduction electrons of silicon that is about 40 times as great as that for the conduction electrons of copper. Why is silicon not a better conductor of electricity than copper?
19. Consider these two statements: (*a*) At low enough temperatures silicon ceases to be a semiconductor and becomes a rather good insulator. (*b*) At high enough temperatures silicon ceases to become a semiconductor and becomes a rather good conductor. Discuss the extent to which each statement is either true or not true.
20. Does the electrical conductivity of an intrinsic (undoped) semiconductor depend on the temperature? On the energy gap E_g between the full and empty bands?
21. How do you account for the fact that the resistivity of metals increases with temperature but that of semiconductors decreases?
22. The energy gaps for the semiconductors silicon and germanium are 1.1 eV and 0.67 eV, respectively. Which substance do you expect would have the higher density of charge carriers at room temperature? At the absolute zero temperature?
23. Discuss this sentence: "The distinction between a metal and a semiconductor is sharp and clear-cut, but that between a semiconductor and an insulator is not."
24. The Hall effect is much greater in semiconductors than in metals. Why? What practical use can be made of this result?
25. Does a slab of *n*-type material carry a *net* negative charge?
26. Suppose that a semiconductor contains equal numbers of donor and acceptor impurities. Do they cancel each other in their electrical effects? If so, what is the mechanism? If not, why not?
27. Why does an *n*-type semiconductor have so many more electrons than holes? Why does a *p*-type semiconductor have so many more holes than electrons? Explain in your own words.
28. What elements other than phosphorus are good candidates to use as donor impurities in silicon? What elements other than aluminum are good candidates to use as acceptor impurities? Consult the periodic table.
29. Does one distinguish between majority and minority carriers for an intrinsic semiconductor such as silicon or germanium? If not, why not? If so, what criterion do you use?
30. In preparing *n*-type or *p*-type semiconductors by doping, why is it extremely important to avoid contamination of the sample with even very small concentrations of unwanted impurities?
31. Would you expect doping to change the resistivity of silicon by very much?
32. When a current flows through a *p*-type material, positive holes move toward the negative terminal of the battery and

combine with electrons in the ohmic electrode connected to the boundary of the crystal. Why doesn't the crystal become negatively charged?

33. Why is silicon often preferred to germanium for making semiconductor devices?

34. Germanium and silicon are similar semiconducting materials whose principal distinction is that the gap width E_g is 0.67 eV for the former and 1.1 eV for the latter. If you wished to construct a *pn* junction (see Fig. 49-10) in which the back current is to be kept as small as possible, which material would you choose and why?

35. In a *pn* junction (see Fig. 49-10) we have seen that electrons and holes may diffuse, in opposite directions, through the junction region. What is the eventual fate of each such particle as it diffuses into the material on the opposite side of the junction? Why is it that the electrons and positive holes do not *all* recombine, thus removing the possibility of conduction?

36. Consider two possible techniques for fabricating a *pn* junction (see Fig. 49-10). (*a*) Prepare separately an *n*-type and a *p*-type sample and join them together, making sure that their abutting surfaces are planar and highly polished. (*b*) Prepare a single *n*-type sample and diffuse an excess acceptor impurity into it from one face, at high temperature. Which method is preferable and why?

37. The *pn* junction shown in Fig. 49-11*a* has equal dopant concentrations on each side of its junction plane. Suppose, however, that the donor concentration were significantly greater than the acceptor concentration. Would the depletion zone still be symmetrically located about the junction plane? If not, would the central plane of the zone move toward the *n*-type or toward the *p*-type face of the junction? Give your reason.

38. Why can't you measure the contact potential difference generated at a *pn* junction by simply connecting a voltmeter across it?

39. In Fig. 49-10*b*, why does the depletion zone build up close to the junction plane? Why does it not spread out throughout the volume of the sample?

40. What does it mean to say that a *pn* junction is biased in the forward direction?

41. (*a*) Discuss the motions of the majority carriers (both electrons and holes) in a forward-biased *pn* junction. (*b*) Discuss the motions of the minority carriers in this same junction.

42. Explain in your own words how the thickness of the depletion zone of a *pn* junction can be decreased by (*a*) increasing the forward-bias voltage and (*b*) increasing the dopant concentration.

43. If you increase the temperature of a reverse-biased *pn* junction, what happens to the current (see Fig. 49-14*a*)? Is the effect larger for silicon or for germanium? (The intrinsic energy gap E_g for silicon is larger than that for germanium.)

44. Does the diode rectifier whose characteristics are shown in Fig. 49-12 obey Ohm's law? What is your criterion for deciding?

45. We have seen that a simple intrinsic (undoped) semiconductor cannot be used as a light-emitting diode. Why not? Would a heavily doped *n*-type or *p*-type semiconductor work?

46. Section 49-9 discussed the mode of operation of a light-emitting diode, in which light is emitted when charge carriers are injected across the central plane of a *pn* junction by an external potential. The reverse device, a *photodiode,* is also a possibility. That is, you can shine light on a *pn* junction and a current will develop across the junction plane. Discuss how such a device might operate. Would it be best to operate it in a forward- or a reverse-biased mode?

47. Explain in your own words how the MOSFET device of Fig. 49-21 works.

48. Do you think that there is a correlation between the critical temperature of a superconductor (Table 49-3) and its electrical conductivity (inverse of resistivity) at room temperature?

EXERCISES

49-1 Quantum Theory of Solids

49-2 Conduction Electrons in a Metal

1. (*a*) Show that Eq. 49-2 can be written as

$$n(E) = CE^{1/2},$$

where $C = 6.81 \times 10^{27}\ \text{m}^{-3}\cdot\text{eV}^{-3/2}$. (*b*) Use this relation to verify a calculation of Sample Problem 49-1—namely, that for $E = 5.00$ eV, $n(E) = 1.52 \times 10^{28}\ \text{m}^{-3}\cdot\text{eV}^{-1}$.

2. Calculate the density $n(E)$ of conduction electron states in a metal for $E = 8.00$ eV and show that your result is consistent with the curve of Fig. 49-1*c*.

3. Gold is a monovalent metal with a molar mass of 197 g/mol and a density of 19.3 g/cm^3 (see Appendix D). Calculate the density of charge carriers.

4. At what pressure would an ideal gas have a density of molecules equal to that of the density of the conduction electrons in copper ($= 8.49 \times 10^{28}\ \text{m}^{-3}$)? Assume that $T = 297$ K.

5. The density and molar mass of sodium are 971 kg/m^3 and 23.0 g/mol, respectively; the radius of the ion Na$^+$ is 98 pm. (*a*) What fraction of the volume of metallic sodium is available to its conduction electrons? (*b*) Carry out the same calculation for copper. Its density, molar mass, and ionic radius are, respectively, 8960 kg/m^3, 63.5 g/mol, and 96 pm. (*c*) For which of these two metals do you think the conduction electrons behave more like a free-electron gas?

49-3 Filling the Allowed States

6. Calculate the probability that a state 0.0730 eV above the Fermi energy is occupied at (*a*) $T = 0$ K and (*b*) $T = 320$ K.

7. The Fermi energy of silver is 5.5 eV. (*a*) At $T = 0°$C, what are the probabilities that states at the following energies are occupied: 4.4 eV, 5.4 eV, 5.5 eV, 5.6 eV, 6.4 eV? (*b*) At what temperature will the occupation probability be 0.16 for a state at 5.6 eV?

8. Prove that the occupancy probabilities for two states whose energies are equally spaced above and below the Fermi energy add up to one.

9. The density of gold is 19.3 g/cm^3. Each atom contributes one conduction electron. Calculate the Fermi energy of gold. See Appendix D for the molar mass of gold.

10. Figure 49-2*c* shows the density of occupied states $n_o(E)$ of the conduction electrons in copper at 1000 K. Calculate $n_o(E)$ for copper for the energies $E = 4.00$, 6.75, 7.00, 7.25, and 9.00 eV. The Fermi energy of copper is 7.06 eV.

11. In Section 47-2 we considered the situation of an electron trapped in an infinitely deep well. Suppose that 100 electrons are placed in a well of width 120 pm, two to a level with opposite spins. Calculate the Fermi energy of the system. (*Note:* The Fermi energy is the energy of the highest occupied level at the absolute zero of temperature.)

12. The conduction electrons in a metal behave like an ideal gas if the temperature is high enough. In particular, the temperature must be such that $kT \gg E_F$, the Fermi energy. What temperatures are required for copper ($E_F = 7.06$ eV) to satisfy this requirement? Compare your answer with the boiling point of copper; see Appendix D. Study Fig. 49-2*c* in this connection and note that we have $kT \ll E_F$ for the conditions of that figure. This is just the reverse of the requirement cited above.

13. Show that Eq. 49-5 can be written as

$$E_F = An^{2/3},$$

where the constant A has the value $3.65 \times 10^{-19}\ \text{m}^2 \cdot \text{eV}$.

14. The Fermi energy of copper is 7.06 eV. (*a*) For copper at 1050 K, find the energy at which the occupancy probability is 0.910. For this energy, evaluate (*b*) the density of states and (*c*) the density of occupied states.

15. Show that the density of states function given by Eq. 49-2 can be written in the form

$$n(E) = \tfrac{3}{2} n E_F^{-3/2} E^{1/2}.$$

Explain how it can be that $n(E)$ is independent of material when the Fermi energy E_F (= 7.06 eV for copper, 9.44 eV for zinc, etc.) appears explicitly in this expression.

16. Show that the probability p_h that a *hole* exists at a state of energy E is given by

$$p_h = \frac{1}{e^{-(E-E_F)/kT} + 1}.$$

(*Hint:* The existence of a hole means that the state is unoccupied; convince yourself that this implies that $p_h = 1 - p$.)

17. White dwarf stars represent a late stage in the evolution of stars like the Sun. They become dense enough and hot enough that we can analyze their structure as a solid in which all Z electrons per atom are free. For a white dwarf with a mass equal to that of the Sun and a radius equal to that of the Earth, calculate the Fermi energy of the electrons. Assume the atomic structure to be represented by iron atoms, and $T = 0$ K.

18. A neutron star can be analyzed by techniques similar to those used for ordinary metals. In this case the neutrons (rather than electrons) obey the probability function, Eq. 49-6. Consider a neutron star of 2.00 solar masses with a radius of 10.0 km. Calculate the Fermi energy of the neutrons.

19. Estimate the number N of conduction electrons in a metal that have energies greater than the Fermi energy as follows. Strictly, N is given by

$$N = \int_{E_F}^{\infty} n(E)p(E)dE.$$

By studying Fig. 49-2*c*, convince yourself that, to a good degree of approximation, this expression can be written as

$$N = \int_{E_F}^{E_F+4kT} n(E_F)(\tfrac{1}{4})dE.$$

By substituting the density of states function, evaluated at the Fermi energy, show that this yields, for the fraction f of conduction electrons excited to energies greater than the Fermi energy,

$$f = \frac{N}{n} = \frac{3kT/2}{E_F}.$$

Why not evaluate the first integral above directly without resorting to an approximation?

20. Use the result of Exercise 19 to calculate the fraction of excited electrons in copper at temperatures of (*a*) absolute zero, (*b*) 300 K, and (*c*) 1000 K.

21. At what temperature will the fraction of excited electrons in lithium equal 0.0130? The Fermi energy of lithium is 4.71 eV. See Exercise 19.

22. Silver melts at 962°C. At the melting point, what fraction of the conduction electrons are in states with energies greater than the Fermi energy of 5.5 eV? See Exercise 19.

49-4 Electrical Conduction in Metals

23. Silver is a monovalent metal. Calculate (*a*) the number of conduction electrons per cubic meter, (*b*) the Fermi energy, (*c*) the Fermi speed, and (*d*) the de Broglie wavelength corresponding to this speed. Extract needed data from Appendix D.

24. Zinc is a bivalent metal. Calculate (*a*) the number of conduction electrons per cubic meter, (*b*) the Fermi energy, (*c*) the Fermi speed, and (*d*) the de Broglie wavelength corresponding to this speed. See Appendix D for needed data on zinc.

25. For silver, calculate (*a*) the mean free path of conduction electrons and (*b*) the ratio of the mean free path to the distance between neighboring ion cores. Silver has a Fermi energy of 5.51 eV and a resistivity of $1.62 \times 10^{-8}\ \Omega \cdot \text{m}$. See Exercise 23.

49-5 Bands and Gaps

49-6 Conductors, Insulators, and Semiconductors

26. Repeat the calculation of Sample Problem 49-6 for a temperature of (*a*) 1000 K and (*b*) 4.0 K.

27. The Fermi–Dirac distribution function can be applied to semiconductors as well as to metals. In semiconductors, E is the energy above the top of the valence band. The Fermi level for an intrinsic semiconductor is nearly midway between the top of the valence band and the bottom of the conduction band. For germanium these bands are separated by a gap of

0.67 eV. Calculate the probability that (*a*) a state at the bottom of the conduction band is occupied and (*b*) a state at the top of the valence band is unoccupied at 290 K.

28. The band gap in pure germanium is 0.67 eV. Assume that the Fermi level is at the middle of the gap. (*a*) Calculate the probability that a state at the bottom of the conduction band is occupied at 16°C. (*b*) At what temperature will the occupation probability of this state be 3.0 times the probability at 16°C?

49-7 Doped Semiconductors

29. Pure silicon at 300 K has an electron density in the conduction band of $1.5 \times 10^{16}\ \mathrm{m}^{-3}$ and an equal density of holes in the valence band. Suppose that one of every 1.0×10^7 silicon atoms is replaced by a phosphorus atom. (*a*) What charge carrier density will the phosphorus add? Assume that all the donor electrons are in the conduction band. (See Appendix D for needed data on silicon.) (*b*) Find the ratio of the charge carrier density in the doped silicon to that for the pure silicon.

30. What mass of phosphorus would be needed to dope a 1.0-g sample of silicon to the extent described in Sample Problem 49-7?

31. A silicon crystal is doped with phosphorus to a concentration of 10^{22} phosphorus atoms per cubic meter. On average, how far apart are these atoms? See Sample Problem 49-7.

32. A sample of very pure germanium has one impurity atom to 1.3×10^9 atoms of germanium. Calculate the distance between impurity atoms.

33. In Fig. 49-24, two energy bands of a hypothetical solid are represented. The bands are filled to level E_x, which may be in either band 1 or band 2. There may be an impurity level at E_i. Indicate whether the solid is a conductor, insulator, intrinsic semiconductor, or extrinsic semiconductor. The impurity type may be donor, acceptor, or none, and extrinsic semiconductors may be either *p*-type or *n*-type. Complete the table.

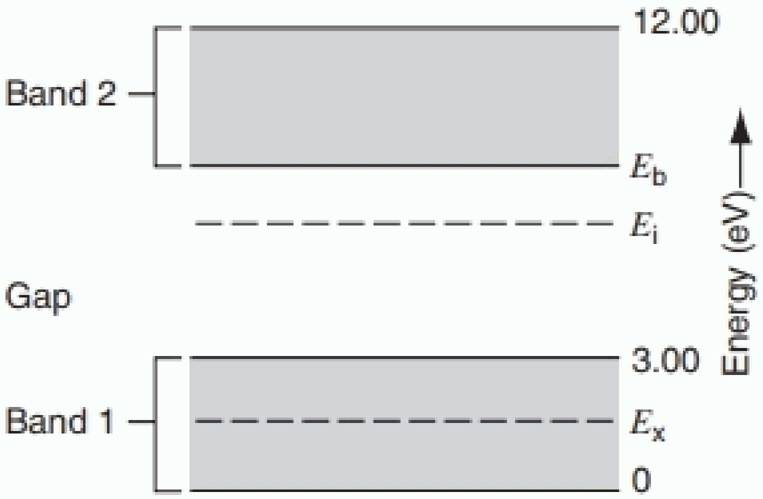


FIGURE 49-24. Exercise 33.

E_x (eV)	E_i (eV)	E_b (eV)	*Type*		
			Solid	*Impurity*	*Extrinsic Semiconductor*
3.00	—	9.00			
3.00	4.06	4.10			
3.00	—	4.10			
1.49	—	9.00			
4.40	—	4.10			
3.00	3.04	4.10			

49-8 The *pn* Junction

34. When a photon enters the depletion region of a *pn* junction, electron–hole pairs can be created as electrons absorb part of the photon's energy and are excited from the valence band to the conduction band. These junctions are thus often used as detectors for photons, especially for x rays and nuclear gamma rays. When a 662-keV gamma-ray photon is totally absorbed by a semiconductor with an energy gap of 1.1 eV, on the average how many electron–hole pairs are created?

35. Calculate and compare the resistances of the diode rectifier for the two points shown on the characteristic curve of Fig. 49-12. The current for the left-hand dot (too small to show in the figure) is 50 pA.

36. (*a*) A capacitance is associated with a *pn* junction. Explain why. (*b*) Derive an expression for the capacitance of the *pn* junction of Sample Problem 49-9.

49-9 Optical Electronics

37. (*a*) Calculate the maximum wavelength that will produce photoconduction in diamond, which has a band gap of 5.5 eV. (*b*) In what part of the electromagnetic spectrum does this wavelength lie?

38. In a particular crystal, the highest occupied band of states is full. The crystal is transparent to light of wavelengths longer than 295 nm but opaque at shorter wavelengths. Calculate the width, in electron-volts, of the gap between the highest occupied band and the next (empty) band.

39. The KCl crystal has a band gap of 7.6 eV above the topmost occupied band, which is full. Is this crystal opaque or transparent to radiation of wavelength 140 nm?

40. (*a*) Fill in the seven-segment display shown in Fig. 49-15*a* to show how all ten numbers may be generated. (*b*) If the numbers are displayed randomly, in what fraction of the displays will each of the seven segments be used?

49-10 The Transistor

49-11 Superconductors

PROBLEMS

1. The Fermi energy of aluminum is 11.66 eV; its density is 2.70 g/cm³ and its molar mass is 27.0 g/mol (see Appendix D). From these data, determine the number of free electrons per atom.

2. Show that, at the absolute zero of temperature, the average energy E_{av} of the conduction electrons in a metal is equal to $\frac{3}{5}E_F$, where E_F is the Fermi energy. [*Hint:* Note that, by definition of average, $E_{\mathrm{av}} = (1/n)\int E n_o(E)\,dE$.]

3. (*a*) Using the result of Problem 2, estimate how much energy would be released by the conduction electrons in a penny (assumed all copper; mass = 3.1 g) if we could suddenly turn off the Pauli exclusion principle. (*b*) For how long would this amount of energy light a 100-W lamp? Note that there is no known way to turn off the Pauli principle!

4. In a simplified model of an intrinsic semiconductor (no doping), the actual distribution in energy of states is replaced by one in which there are N_v states in the valence band, all these states having the same energy E_v, and N_c states in the conduction band, all these states having the same energy E_c. The number of electrons in the conduction band equals the number of holes in the valence band. (*a*) Show that this last condition implies that

$$\frac{N_c}{e^{(E_c - E_F)/kT} + 1} = \frac{N_v}{e^{-(E_v - E_F)/kT} + 1}.$$

(*Hint:* See Exercise 16.) (*b*) If the Fermi level is in the gap between the two bands and is far from both bands compared to kT, then the exponentials dominate in the denominators. Under these conditions, show that

$$E_F = \tfrac{1}{2}(E_c + E_v) + \tfrac{1}{2}kT\ln(N_v/N_c),$$

and therefore that, if $N_v \approx N_c$, the Fermi level is close to the center of the gap.

5. Doping changes the Fermi energy of a semiconductor. Consider silicon, with a gap of 1.1 eV between the valence and conduction bands. At 290 K, the Fermi level of the pure material is nearly at the midpoint of the gap. Suppose that it is doped with donor atoms, each of which has a state 0.15 eV below the bottom of the conduction band, and suppose further that doping raises the Fermi level to 0.084 eV below the bottom of that band. (*a*) For both the pure and doped silicon, calculate the probability that a state at the bottom of the conduction band is occupied. (*b*) Also calculate the probability that a donor state in the doped material is occupied. See Fig. 49-25.

6. A silicon sample is doped with atoms having a donor state 0.11 eV below the bottom of the conduction band. (*a*) If each of these states is occupied with probability 4.8×10^{-5} at temperature 290 K, where is the Fermi level relative to the top of the valence band? (*b*) What then is the probability that a state at the bottom of the conduction band is occupied? The energy gap in silicon is 1.1 eV.

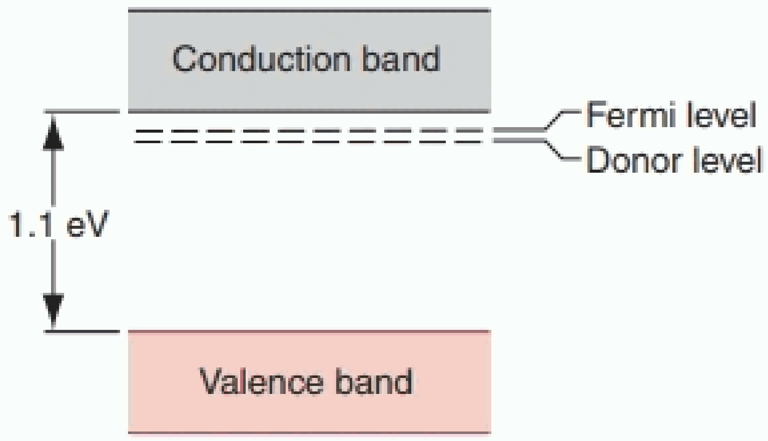


FIGURE 49-25. Problem 5.

7. For an ideal *pn*-junction diode, with a sharp boundary between the two semiconducting materials, the current i is related to the potential difference ΔV across the diode by

$$i = i_0(e^{e\Delta V/kT} - 1),$$

where i_0, which depends on the materials but not on the current or potential difference, is called the *reverse saturation current.* ΔV is positive if the junction is forward biased and negative if it is reverse biased. (*a*) Verify that this expression predicts the behavior expected of a diode by sketching i as a function of ΔV over the range $-0.12\text{ V} < \Delta V < +0.12\text{ V}$. Take $T = 290$ K and $i_0 = 5.0$ nA. (*b*) For the same temperature, calculate the ratio of the current for a 0.50-V forward bias to the current for a 0.50-V reverse bias.

8. A drop of lead (work function = 3.4 eV) is in close contact with a sheet of copper (work function = 4.5 eV). Find the contact potential difference that appears across the lead–copper interface. How might you measure it? Draw an energy diagram, showing (in the style of Fig. 49-4*b*) the relative Fermi levels both before and after the two metals are joined together. Can such a junction serve as a diode rectifier?

CHAPTER 50

NUCLEAR PHYSICS

Deep within the atom lies its nucleus, occupying only 10^{-15} of the volume of the atom but providing most of its mass as well as the force that holds it together. The next goal in our study of physics is to understand the structure of the nucleus and the substructure of its components.

Our task is made easier by the many similarities between the study of atoms and the study of nuclei. Both systems are governed by the laws of quantum mechanics. Like atoms, nuclei have excited states that can decay to the ground state through the emission of photons (gamma rays). In certain circumstances, as we shall see, nuclei can exhibit shell effects that are very similar to those of atoms. We shall also see that there are differences between the study of atoms and the study of nuclei that keep us from achieving as complete an understanding of nuclei as we have of atoms.

In this chapter we study the structure of nuclei and their constituents. We consider some experimental techniques for studying their properties, and we conclude with a description of the theoretical basis for understanding the structure of nuclei.

50-1 DISCOVERING THE NUCLEUS

In the first years of the 20th century, not much was known about the structure of atoms beyond the fact that they contained electrons. This particle had been discovered (by J. J. Thomson) only in 1897, and its mass was unknown in those early days. Thus it was not possible even to say just how many electrons a given atom contained. Atoms are electrically neutral so they must also contain some positive charge, but at that time nobody knew what form this compensating positive charge took. How the electrons moved within the atom and how the mass of the atom was divided between the electrons and the positive charge were also open questions.

In 1911, Ernest Rutherford, interpreting some experiments carried out in his laboratory, was led to propose that the positive charge of the atom was densely concentrated at the center of the atom and that, furthermore, it was responsible for most of the mass of the atom. He had discovered the atomic nucleus!

Until this step had been taken, all attempts to understand the motions of the electrons within the atom were doomed to failure. Only 2 years after Rutherford's proposal, Niels Bohr used the concept of the nuclear atom to develop the semiclassical theory of atomic structure that we described in Chapter 47. This early work by Rutherford and Bohr marks the beginning of our understanding of the structure of atoms.

How did Rutherford come to make this proposal? It was not an idle conjecture but was based firmly on the results of an experiment suggested by him and carried out by his collaborators, Hans Geiger (of Geiger counter fame) and Ernest Marsden, a 20-year-old student who had not yet earned his bachelor's degree.

Rutherford's idea was to probe the forces acting within an atom by firing energetic alpha (α) particles through a thin target foil and measuring the extent to which they were deflected as they passed through the foil. Alpha particles, which are about 7300 times more massive than electrons, carry a charge of $+2e$ and are emitted spontaneously (with energies of a few MeV) by many radioactive materials. We now know that these useful projectiles are the nuclei of the atoms of ordinary helium. Figure 50-1 shows the experimental arrangement of Geiger and Marsden. The experiment consists of counting the number of α particles deflected through various scattering angles θ. (See Section 26-8.)

Figure 50-2 shows their results. Note especially that the vertical scale is logarithmic. We see that most of the α particles are scattered through rather small angles, but—and this was the big surprise—some of them are scattered through very large angles, approaching 180°. In Rutherford's words: "It was quite the most incredible event that ever happened to me in my life. It was almost as incredible as if you had fired a 15-inch shell at a piece of tissue paper and it came back and hit you."

Why was Rutherford so surprised? At the time of these experiments, many physicists believed in a model of the atom that had been proposed by J. J. Thomson. In Thomson's model, the positive charge of the atom was thought to be spread out through the entire volume of the atom. The electrons were thought to be distributed throughout this volume, somewhat like seeds in a watermelon, and to vibrate about their equilibrium positions within this sphere of charge.

The maximum deflecting force acting on the α particle as it passes through such a positive sphere of charge proves to be far too small to deflect the α particle by even as much as one degree. The electrons in the atom would also have very little effect on the massive, energetic α particle. They would, in fact, be themselves strongly deflected, much as a swarm of gnats would be brushed aside by a stone thrown through them. There is simply no mechanism in Thomson's atom model to account for the backward deflection of an α particle.

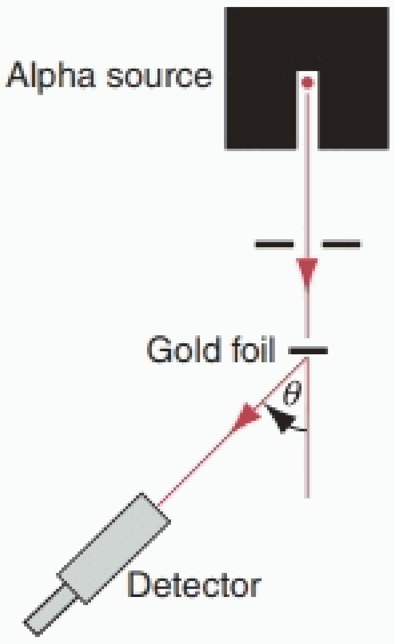


FIGURE 50-1. The experimental arrangement used in Rutherford's laboratory to study the scattering of α particles by thin metal foils. The detector can be rotated to various scattering angles θ.

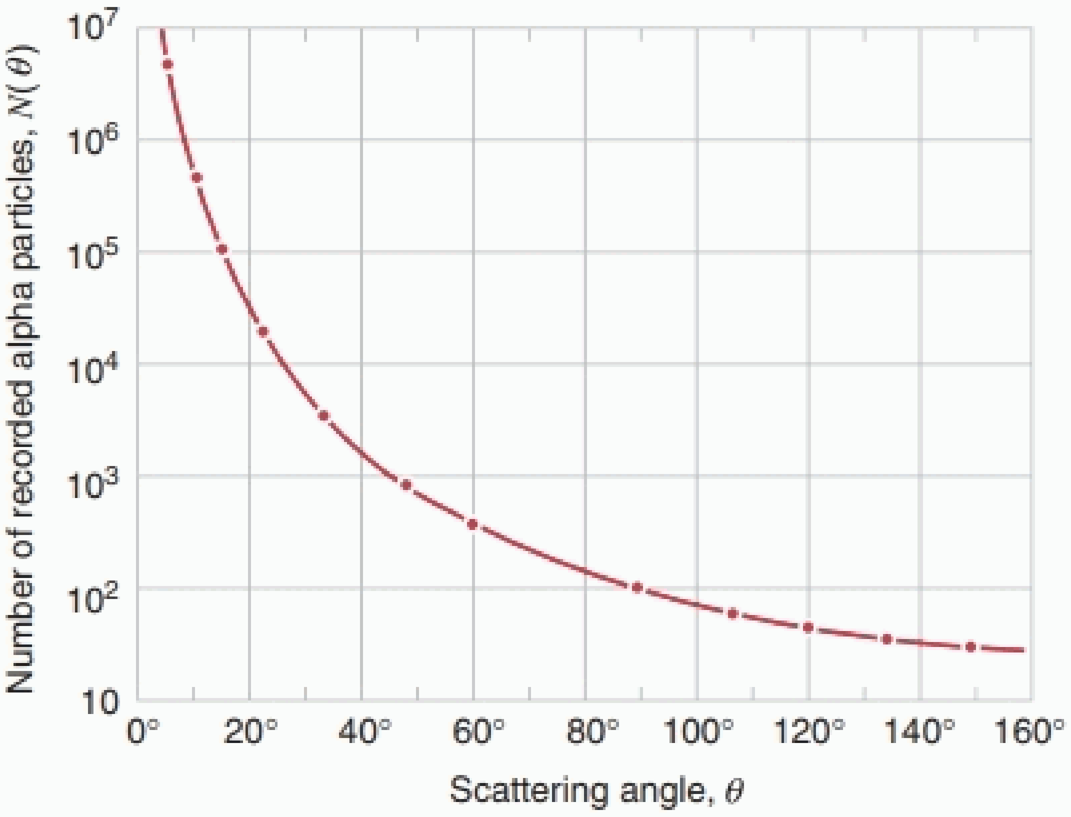


FIGURE 50-2. The dots show the α-particle scattering results from the experiments of Geiger and Marsden, and the solid curve is computed according to Rutherford's theory of the nucleus. Note that the vertical axis is marked in powers of 10.

Rutherford saw that to produce such a large deflection there must be a large force, which could be provided if the positive charge were concentrated tightly at the center of the atom, instead of being spread throughout its volume. On this model the incoming α particle can get very close to the center of the positive charge without penetrating it, resulting in a large deflecting force; see Sample Problem 50-1.

Figure 50-3 shows the paths taken by typical α particles as they pass through the atoms of the target foil. As we see, most are deflected only slightly or not at all, but a few (those whose extended incoming paths pass, by chance, close to a nucleus) are deflected through large angles. From an analysis of the data, Rutherford concluded that the dimensions of the nucleus must be smaller than the diameter

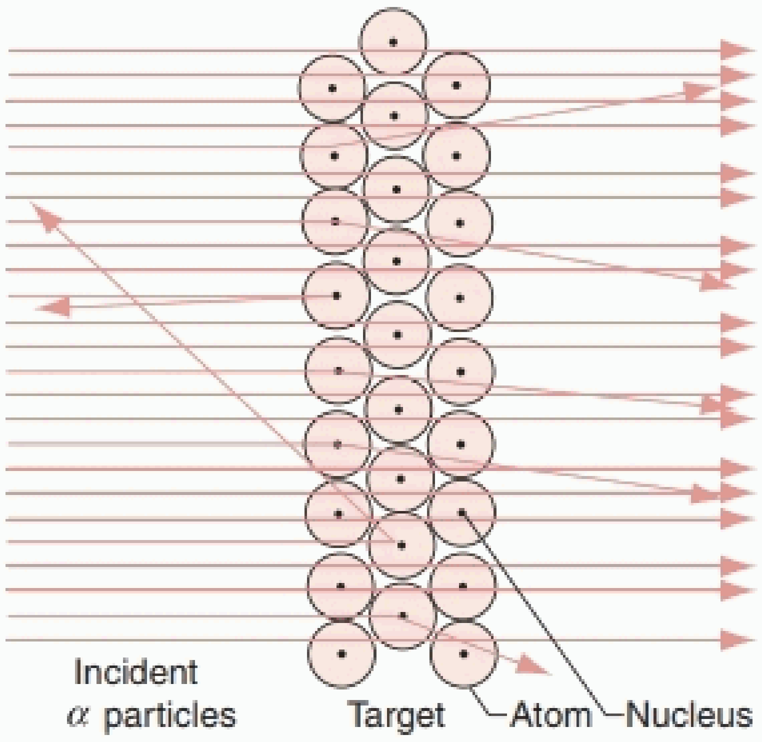


FIGURE 50-3. The angle through which an α particle is scattered depends on how close its extended incident path lies to the nucleus of an atom. Large deflections result only from very close encounters.

of an atom by a factor of about 10^4. The atom is mostly empty space! It is not often that the piercing insight of a gifted scientist, supported by a few simple calculations,* leads to results of such importance.

SAMPLE PROBLEM 50-1. A 5.30-MeV α particle happens, by chance, to be headed directly toward the nucleus of an atom of gold ($Z = 79$). How close does it get before it comes momentarily to rest and reverses its course? Neglect the recoil of the (relatively massive) gold nucleus.

Solution Initially the total mechanical energy of the two interacting particles is just equal to K_α ($= 5.30$ MeV), the initial kinetic energy of the α particle. The potential energy U is taken to be zero when the particles are separated by a large distance. At the moment the α particle comes to rest, the total energy is the electrostatic potential energy of the system of two particles (Eq. 28-7, $U = q_1 q_2 / 4\pi\epsilon_0 r$). Because energy must be conserved, these two quantities must be equal, or

$$K_\alpha = \frac{1}{4\pi\epsilon_0}\frac{qQ}{d}$$

in which q ($= 2e$) is the charge of the α particle, Q ($= 79e$) is the charge of the gold nucleus, and d is the distance between the centers of the two particles when the α particle is at rest.

Substituting for the charges and solving for d, we obtain

$$\begin{aligned} d &= \frac{qQ}{4\pi\epsilon_0 K_\alpha} \\ &= (8.99 \times 10^9\ \mathrm{N \cdot m^2/C^2})\frac{(2)(79)(1.60 \times 10^{-19}\ \mathrm{C})^2}{(5.30\ \mathrm{MeV})(1.60 \times 10^{-13}\ \mathrm{J/MeV})} \\ &= 4.29 \times 10^{-14}\ \mathrm{m} = 42.9\ \mathrm{fm}. \end{aligned}$$

This is a small distance by atomic standards but not by nuclear standards. As we shall see in the following section, it is considerably larger than the sum of the radii of the gold nucleus and the α particle. Thus the α particle reverses its course without ever "touching" the gold nucleus.

If the positive charge associated with the gold atom had been spread uniformly throughout the volume of the atom, the maximum retarding force acting on the α particle would have occurred at the moment the α particle began to touch the surface of the atom. This force (see Exercise 2) would have been far too weak to have had much effect on the motion of the α particle, which would have gone barreling right through such a "spongy" atom.

50-2 SOME NUCLEAR PROPERTIES

The nucleus, tiny as it may be, has a structure that is every bit as complex as that of the atom. Nuclei are made up of protons and neutrons. These particles (unlike the electron) are not true elementary particles, because they are made up of other particles, called *quarks*. However, nuclear physics—the subject of this chapter—is concerned primarily with studies of the nucleus that do not involve the internal structure of the protons and neutrons themselves. The fundamental nature of these two particles is a topic in the field of elementary particle physics, which we consider in Chapter 52.

Nuclear Systematics

The number of protons in the nucleus (the *proton number*) is often called the atomic number and is represented by Z. The number of neutrons is called the *neutron number*, and we represent it by N. Aside from the difference in their electric charges ($q = +e$ for the proton, $q = 0$ for the neutron), the proton and the neutron are very similar particles: they have nearly equal masses and experience identical nuclear forces inside nuclei. For this reason, we classify the proton and neutron together as *nucleons*. The total number of nucleons ($= Z + N$) is called the *mass number*, and we represent it by A.

By specifying Z and A (and therefore N) we uniquely identify a particular nuclear species or *nuclide*. We use A, the total number of nucleons, as an identifying superscript in labeling nuclides. In ^{81}Br, for example, there are 81 nucleons. The symbol "Br" tells us that we are dealing with bromine, for which $Z = 35$. The remaining 46 nucleons are neutrons, so that, for this nuclide, $Z = 35$, $N = 46$, and $A = 81$. Two nuclides with the same Z but different N and A, such as ^{81}Br and ^{82}Br, are called *isotopes*.

Figure 50-4 shows a chart of the known nuclides as a plot of Z against N. The dark shading represents stable nuclides; the lighter shading represents known radioactive nuclides, or *radionuclides*. Table 50-1 shows some properties of a few selected nuclides.

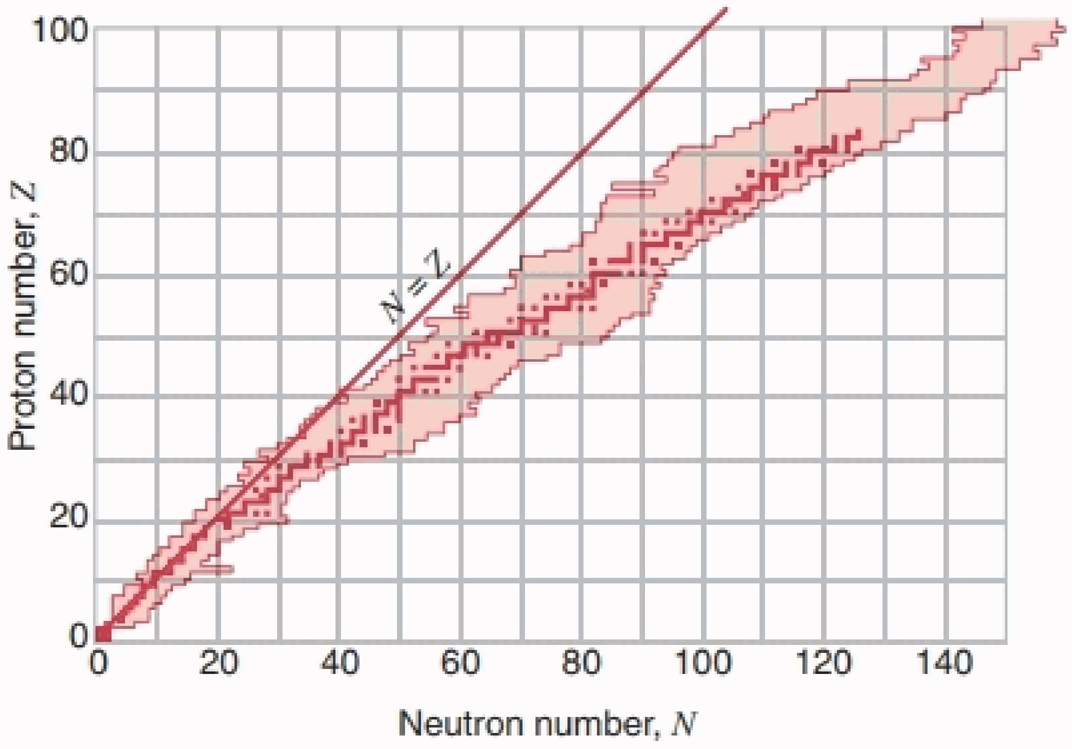


FIGURE 50-4. A plot of the known nuclides. The dark shading indicates stable nuclides and the light shading shows radioactive nuclides. Note that light stable nuclides have essentially equal numbers of protons and neutrons, whereas $N > Z$ for heavy nuclei.

---

*For an analysis of this scattering experiment, see Kenneth S. Krane, *Modern Physics,* 2nd ed. (Wiley, 1996), Chapter 6.

TABLE 50-1 **Some Properties of Selected Nuclides**

Nuclide	*Z*	*N*	*A*	*Stability*[a]	*Atomic Mass* (u)	*Radius* (fm)	*Binding Energy per Nucleon* (MeV)	*Spin* ($h/2\pi$)	*Magnetic Dipole Moment* (μ_N)
^{7}Li	3	4	7	92.5%	7.016004	2.30	5.61	$\frac{3}{2}$	+3.26
^{14}N	7	7	14	99.6%	14.003074	2.89	7.48	1	+0.404
^{31}P	15	16	31	100%	30.973761	3.77	8.48	$\frac{1}{2}$	+1.13
^{88}Rb	37	51	88	18 m	87.911319	5.34	8.68	2	+0.508
^{120}Sn	50	70	120	32.4%	119.902197	5.92	8.50	0	0
^{157}Gd	64	93	157	15.7%	156.923957	6.47	8.20	$\frac{3}{2}$	−0.340
^{197}Au	79	118	197	100%	196.966552	6.98	7.92	$\frac{3}{2}$	+0.146
^{239}Pu	94	145	239	24,100 y	239.052157	7.45	7.56	$\frac{1}{2}$	+0.203

[a] For stable nuclides the isotopic abundance is given; for radionuclides, the half-life.

Note that there is a reasonably well-defined zone of stability in Fig. 50-4. Unstable radionuclides lie on either side of the stability zone.

The Nuclear Force

The force that controls the electronic structure and properties of the atom is the familiar Coulomb force. To bind the nucleus together, however, there must be a strong attractive force of a totally new kind acting between the neutrons and the protons. This force must be strong enough to overcome the repulsive Coulomb force between the (positively charged) protons and to bind both neutrons and protons into the tiny nuclear volume. Experiments suggest that this *strong force,* as it is simply called, has the same character between any pair of nuclear constituents, be they neutrons or protons.

The "strong force" has a short range, roughly equal to 10^{-15} m. This means that the attractive force between pairs of nucleons drops rapidly to zero for nucleon separations greater than a certain critical value. This in turn means that, except in the smallest nuclei, a given nucleon cannot interact through the strong force with all the other nucleons in the nucleus but only with a few of its nearest neighbors. By contrast, the Coulomb force is not a short-range force. A given proton in a nucleus exerts a Coulomb repulsion on all the other protons, no matter how large their separation; see Exercise 11.

Figure 50-4 shows that the lightest stable nuclides tend to lie on or close to the line $Z = N$. The heavier stable nuclides lie well below this line and thus typically have many more neutrons than protons. The tendency to an excess of neutrons at large mass numbers is a Coulomb repulsion effect. Because a given nucleon interacts with only a small number of its neighbors through the strong force, the amount of energy tied up in strong-force bonds between nucleons increases in proportion to A. The energy tied up in Coulomb-force bonds between protons increases more rapidly than this because each proton interacts with all other protons in the nucleus. Thus the Coulomb energy becomes increasingly important at high mass numbers.

Consider a nucleus with 238 nucleons. If it were to lie on the $Z = N$ line, it would have $Z = N = 119$. However, such a nucleus, if it could be assembled, would fly apart at once because of Coulomb repulsion. Relative stability is found only if we replace 27 of the protons by neutrons, thus greatly diluting the Coulomb repulsion effect. We then would have the nuclide ^{238}U, which has $Z = 92$ and $N = 146$, a neutron excess of 54.

Even in ^{238}U, Coulomb effects are evident in that (1) this nuclide is radioactive and emits α particles, and (2) it can easily break up (fission) into two fragments. Both of these processes reduce the Coulomb energy more than they do the energy in the strong-force bonds.

Nuclear Radii

We have used the Bohr radius a_0 ($= 5.29 \times 10^{-11}$ m $=$ 52.9 pm) as a convenient unit for measuring the dimensions of atoms. Nuclei are smaller by a factor of about 10^4, and a convenient unit for measuring distances of this scale is the *femtometer* ($= 10^{-15}$ m). This unit is often called the *fermi* and shares the same abbreviation. Thus

$$1 \text{ fermi} = 1 \text{ femtometer} = 1 \text{ fm} = 10^{-15} \text{ m}.$$

We can learn about the size and structure of nuclei by doing scattering experiments, much as suggested by Fig. 50-1, using an incident beam of high-energy electrons. The energy of the incident electrons must be high enough (> 200 MeV) so that their de Broglie wavelength will be small enough for them to act as structure-sensitive nuclear probes. In effect, these experiments measure the diffraction pattern of the scattered particles and so deduce the shape of the scattering object (the nucleus).

From a variety of scattering experiments, the nuclear density has been deduced to be of the form shown in Fig. 50-5. We see that the nucleus does not have a sharply defined surface. It does, however, have a characteristic mean

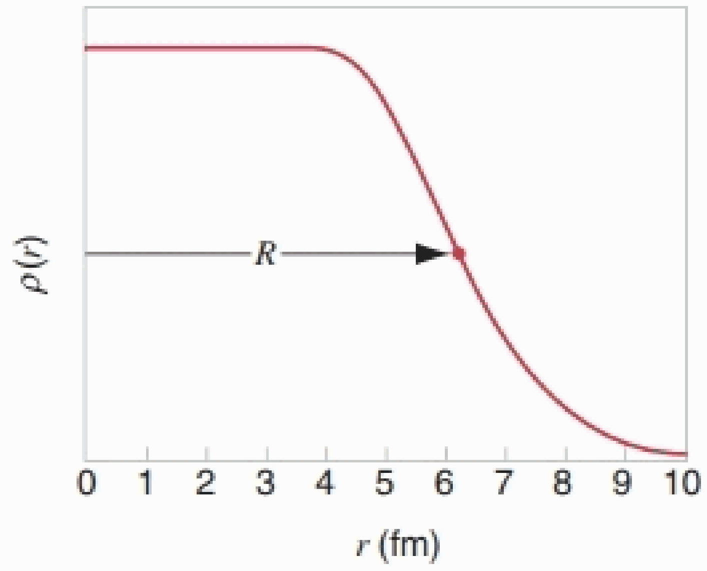


FIGURE 50-5. The variation with radial distance of the density of a nucleus of ^{197}Au.

radius R. The density $\rho(r)$ has a constant value in the nuclear interior and drops to zero through the fuzzy surface zone. From these experiments it has been found that R increases with A approximately as

$$R = R_0 A^{1/3}, \tag{50-1}$$

in which A is the mass number and R_0 is a constant with a value of about 1.2 fm. For ^{63}Cu, for example,

$$R = (1.2\text{ fm})(63)^{1/3} = 4.8\text{ fm}.$$

By comparison, the mean radius of a copper ion in a lattice of solid copper is 1.8 Bohr radii, about 2×10^4 times larger.

Nuclear Masses and Binding Energies

Atomic masses can be measured with great precision using modern mass spectrometer and nuclear reaction techniques. We recall that such masses are measured in *unified atomic mass units* (abbreviation u), chosen so that the atomic mass (*not* the nuclear mass) of ^{12}C is exactly 12 u. The relation of this unit to the SI mass standard is, to six significant figures,

$$1\text{ u} = 1.66054 \times 10^{-27}\text{ kg}.$$

Note that the mass number (symbol A) identifying a nuclide is so named because this number is equal to the atomic mass of the nuclide, rounded to the nearest integer. Thus the mass number of the nuclide ^{137}Cs is 137. This nuclide contains 55 protons and 82 neutrons, a total of 137 particles; its atomic mass is 136.907084 u, which rounds off numerically to 137.

In nuclear physics, as contrasted with atomic physics, the energy changes per event are commonly so great that Einstein's well-known mass–energy relation $E = \Delta mc^2$ is an indispensable workaday tool. We shall often need to use the energy equivalent of 1 atomic mass unit, and we find it (using constants evaluated to six significant figures) from

$$E = \Delta m\, c^2 = \frac{(1.66054 \times 10^{-27}\text{ kg})(2.99792 \times 10^8\text{ m/s})^2}{1.60218 \times 10^{-13}\text{ J/MeV}}$$
$$= 931.5\text{ MeV}.$$

This means that we can write c^2 as 931.5 MeV/u and can thus easily find the energy equivalent (in MeV) of any mass or mass difference (in u), or conversely.

As an example, consider the deuteron, the nucleus of the heavy hydrogen atom. A deuteron consists of a proton and a neutron bound together by the strong force. The energy E_B that we must add to the deuteron to tear it apart into its two constituent nucleons is called its *binding energy.* In effect, the binding energy is a measure of the total internal energy of the nucleus, due in part to the strong force between the nucleons, the Coulomb force between the nucleons, and the kinetic energies of the nucleons relative to the center of mass of the entire nucleus. From conservation of energy we can write, for this pulling-apart process,

$$m_d c^2 + E_B = m_n c^2 + m_p c^2. \tag{50-2}$$

If we add $m_e c^2$, the energy equivalent of one electron mass, to each side of this equation, we have

$$(m_d + m_e)c^2 + E_B = m_n c^2 + (m_p + m_e)c^2$$

or

$$m(^2\text{H})c^2 + E_B = m_n c^2 + m(^1\text{H})c^2. \tag{50-3}$$

Here $m(^2\text{H})$ and $m(^1\text{H})$ are the masses of the neutral *heavy* hydrogen atom and the neutral *ordinary* hydrogen atom, respectively. They are atomic masses, not nuclear masses. Solving Eq. 50-3 for E_B yields

$$E_B = [m_n + m(^1\text{H}) - m(^2\text{H})]c^2 = \Delta mc^2, \tag{50-4}$$

in which Δm is the mass difference. In making calculations of this kind we always use atomic, rather than nuclear, masses, as this is what is normally tabulated. As in this example, the electron masses conveniently cancel.*

For the deuteron calculation the needed masses are $m_n = 1.008665$ u, $m(^1\text{H}) = 1.007825$ u, and $m(^2\text{H}) = 2.014102$ u. The mass difference in Eq. 50-4 is

$$\Delta m = 1.008665\text{ u} + 1.007825\text{ u} - 2.014102\text{ u}$$
$$= 0.002388\text{ u}.$$

Substituting into Eq. 50-4 and replacing c^2 by its equivalent, 931.5 MeV/u, we find the binding energy to be

$$E_B = (0.002388\text{ u})(931.5\text{ MeV/u}) = 2.224\text{ MeV}.$$

Compare this with the binding energy of the hydrogen atom in its ground state, which is 13.6 eV, about five orders of magnitude smaller.

If we divide the binding energy of a nucleus by its mass number, we get the average binding energy per nucleon, a property we have listed in Table 50-1. Figure 50-6 shows a plot of this quantity as a function of mass number. The fact that this *binding energy curve* "droops" at both high and

---

*See, however, Problem 11, for an exception.

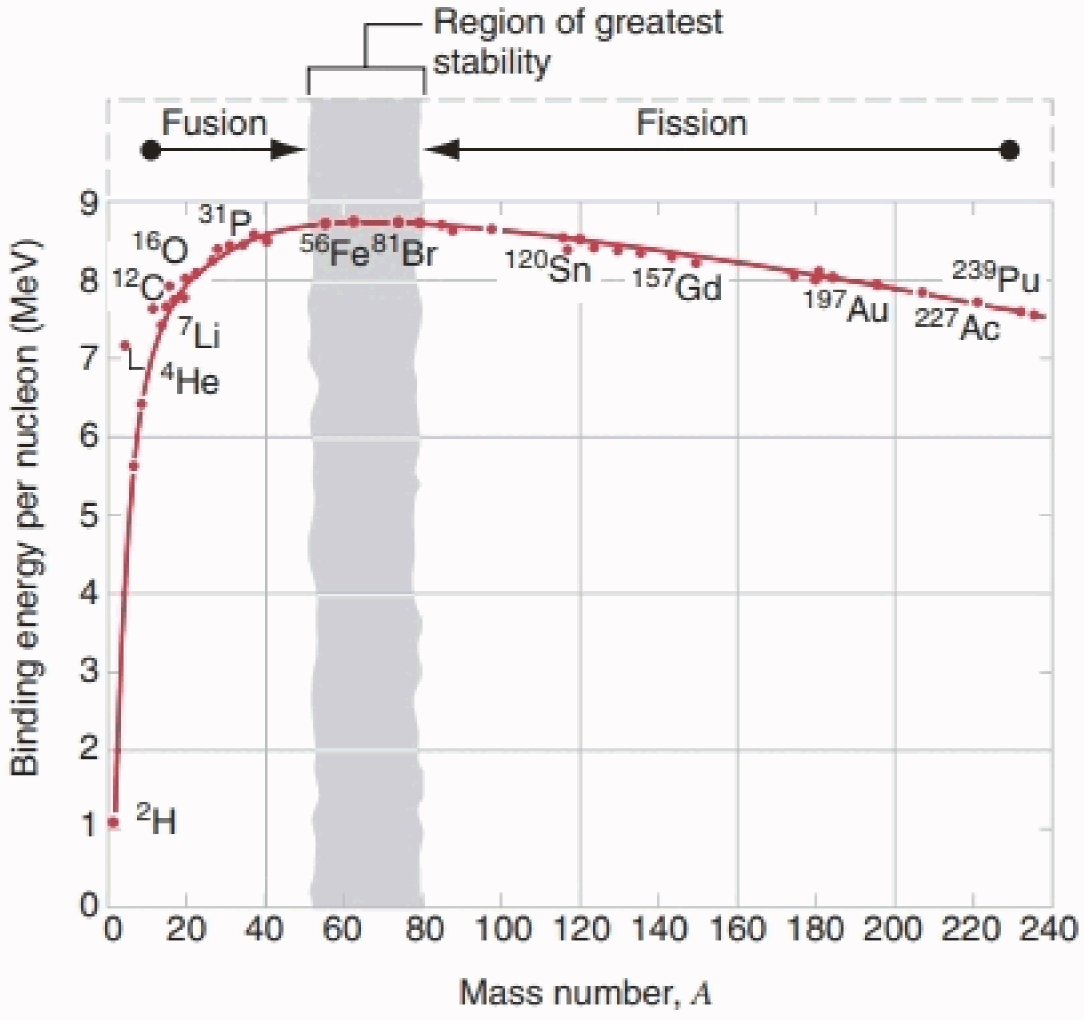


FIGURE 50-6. The binding energy per nucleon over the range of mass numbers. Some of the nuclides of Table 50-1 are identified, along with a few others. The region of greatest stability corresponds to mass numbers from about 50 to 80.

low mass numbers has practical consequences of the greatest importance.*

The drooping of the binding energy curve at high mass numbers tells us that nucleons are more tightly bound when they are assembled into two middle-mass nuclei rather than into a single high-mass nucleus. In other words, energy can be released in the *nuclear fission* of a single massive nucleus into two smaller fragments.

The drooping of the binding energy curve at low mass numbers, on the other hand, tells us that energy will be released if two nuclei of small mass number combine to form a single middle-mass nucleus. This process, the reverse of fission, is called *nuclear fusion.* It occurs inside our Sun and other stars and is the mechanism by which the Sun generates the energy it radiates to us.

Nuclear Spin and Magnetism

Nuclei, like atoms, have an intrinsic angular momentum whose maximum component along any chosen z axis is given by $J(h/2\pi)$. Here J is a quantum number, which may be integral or half-integral, called the *nuclear spin;* some values for selected nuclides are shown in Table 50-1.

Again as for atoms, a nuclear angular momentum has a nuclear magnetic moment associated with it. Recall (see Section 35-3) that in atomic magnetism, the *Bohr magneton* μ_B, defined as

$$\mu_B = \frac{eh}{4\pi m_e} = 5.79 \times 10^{-5}\ \text{eV/T},$$

is a unit of convenience. In nuclear physics the corresponding unit of convenience is the *nuclear magneton* μ_N, defined similarly to the Bohr magneton except that the electron mass m_e is replaced by the proton mass m_p. That is,

$$\mu_N = \frac{eh}{4\pi m_p} = 3.15 \times 10^{-8}\ \text{eV/T}.$$

Because the magnetic moment of the free electron is (very nearly) one Bohr magneton, it might be supposed that the magnetic moment of the free proton would be (very nearly) one nuclear magneton. It is not very close, however, the measured value being $+2.7928\ \mu_N$. To understand the magnetic moments of the proton and neutron, it is necessary to consider their internal structure. The magnetic moments of heavier nuclei can in turn be analyzed in terms of the magnetic moments of the constituent protons and neutrons.

SAMPLE PROBLEM 50-2. What is the approximate density of the *nuclear matter* from which all nuclei are made?

Solution We know that this density is large, because virtually all the mass of the atom resides in its tiny nucleus. The volume of the nucleus, approximated as a uniform sphere of radius R, is given by Eq. 50-1 as

$$V = \tfrac{4}{3}\pi R^3 = \tfrac{4}{3}\pi(R_0^3 A).$$

The density ρ_n of nuclear matter, expressed in nucleons per unit volume, is then

$$\rho_n = \frac{A}{V} = \frac{A}{(4\pi/3)R_0^3 A}$$

$$= \frac{1}{(4\pi/3)(1.2\ \text{fm})^3} = 0.14\ \text{nucleons/fm}^3.$$

The mass of a nucleon is 1.7×10^{-27} kg. The *mass* density ρ_m of nuclear matter is then

$$\begin{aligned}\rho_m &= (0.14\ \text{nucleons/fm}^3)(1.7 \times 10^{-27}\ \text{kg/nucleon}) \\ &\quad \times (1\ \text{fm}/10^{-15}\ \text{m})^3 \\ &= 2.4 \times 10^{17}\ \text{kg/m}^3,\end{aligned}$$

or 2.4×10^{14} times the density of water! Unlike the orbital electrons, the nuclides have a density nearly independent of the number of their nucleons. To some extent nucleons are packed in like marbles in a bag.

SAMPLE PROBLEM 50-3. Imagine that a typical middle-mass nucleus such as ^{120}Sn is picked apart into its constituent protons and neutrons. Find (*a*) the total energy required and (*b*) the energy per nucleon. The atomic mass of ^{120}Sn is 119.902197 u; see Table 50-1.

Solution (*a*) ^{120}Sn contains 50 protons and $120 - 50 = 70$ neutrons. The combined atomic mass of these free particles is

$$\begin{aligned}M &= Zm_p + Nm_n = 50 \times 1.007825\ \text{u} + 70 \times 1.008665\ \text{u} \\ &= 120.997800\ \text{u}.\end{aligned}$$

The Curve of Binding Energy has even been adopted as the title of a book (by John McPhee) about the possibilities of nuclear terrorism!

This exceeds the atomic mass of ^{120}Sn by

$$\Delta m = 120.997800 \text{ u} - 119.902197 \text{ u} = 1.095603 \text{ u}.$$

Converting this to a rest energy yields the total binding energy,

$$E_B = \Delta mc^2 = (1.0956 \text{ u})(931.5 \text{ MeV/u}) = 1020.6 \text{ MeV}.$$

(*b*) The binding energy E per nucleon is

$$E = \frac{E_B}{A} = \frac{1020.6 \text{ MeV}}{120} = 8.50 \text{ MeV/nucleon}.$$

This agrees with the value that may be read from the curve of Fig. 50-6.

50-3 RADIOACTIVE DECAY

As Fig. 50-4 shows, most of the nuclides that have been identified are radioactive. That is, they spontaneously emit a particle, transforming themselves in the process into a different nuclide. In this chapter we discuss the two most common situations: the emission of an α particle (alpha decay) and the emission of an electron (beta decay).

No matter what the nature of the decay, its main feature is that it is statistical. Consider, for example, a 1-mg sample of uranium metal. It contains 2.5×10^{18} atoms of the very long-lived alpha emitter ^{238}U. The nuclei of these atoms have existed without decaying since they were created (before the formation of our solar system) in the explosion of a supernova.

During any given second about 12 of the nuclei in our sample will decay, emitting an α particle in the process. We have absolutely no way of predicting, however, whether any given nucleus in the sample will be among those that do so. Every single ^{238}U nucleus has exactly the same probability as any other to decay during any 1-s observation period—namely, $12/(2.5 \times 10^{18})$, or one chance in 2×10^{17}.

In general, if a sample contains N radioactive nuclei, we can express the statistical character of the decay process by saying that the rate of change of the number of nuclei, or $-dN/dt$ (where the minus sign makes this a positive quantity), is proportional to the number of nuclei present, or

$$-\frac{dN}{dt} = \lambda N, \tag{50-5}$$

in which the proportionality constant λ, called the *disintegration constant,* has a different characteristic value for each radioactive nuclide. We can rewrite Eq. 50-5 as

$$\frac{dN}{N} = -\lambda \, dt,$$

which integrates readily to

$$N = N_0 e^{-\lambda t}. \tag{50-6}$$

Here N_0 is the number of radioactive nuclei in the sample at $t = 0$. We see that the decrease of N with time follows a simple exponential law.

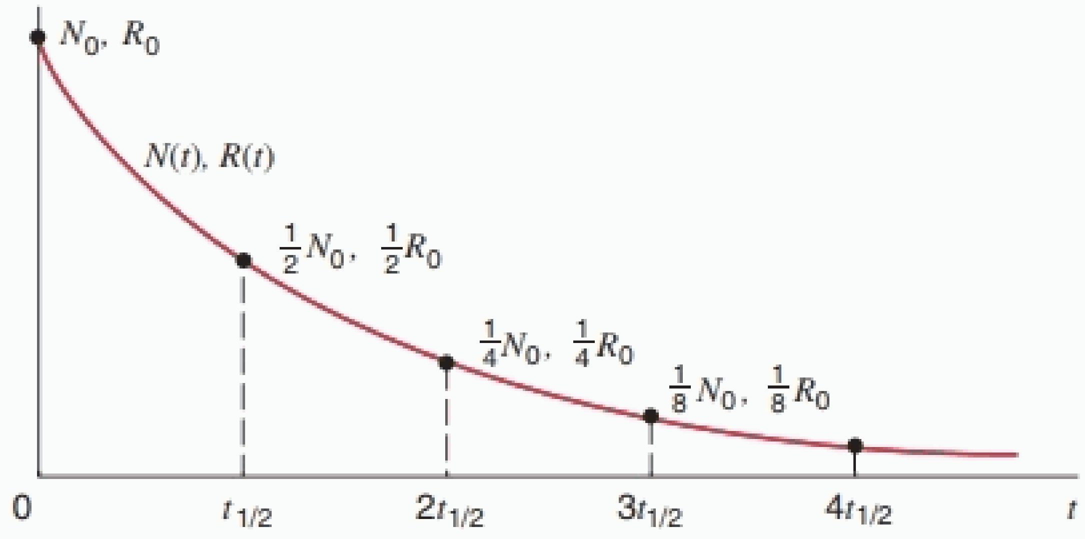


FIGURE 50-7. The exponential decrease in the number of radioactive nuclei and their decay rate.

We are often more interested in the *activity* or decay rate R $(= \lambda N)$ of the sample than we are in N. Multiplying Eq. 50-6 by λ gives

$$R = R_0 e^{-\lambda t}, \tag{50-7}$$

in which R_0 $(= \lambda N_0)$ is the decay rate at $t = 0$. The number of nuclei present and the decay rate both follow the same exponential law.

A quantity of interest is the time $t_{1/2}$, called the *half-life,* after which both N and R are reduced to one-half of their initial values. Putting $R = \frac{1}{2}R_0$ in Eq. 50-7 gives

$$\tfrac{1}{2}R_0 = R_0 e^{-\lambda t_{1/2}},$$

which leads readily to

$$t_{1/2} = \frac{\ln 2}{\lambda}, \tag{50-8}$$

a relationship between the half-life and the disintegration constant. Figure 50-7 shows how both N and R decrease exponentially with time.

The following two sample problems show how λ can be measured for decay processes with relatively short half-lives and also with relatively long half-lives.

SAMPLE PROBLEM 50-4. In short-lived decays, it is possible to measure directly the decrease in the decay rate R with time. The following table gives some data for a sample of ^{128}I, a radionuclide often used medically as a tracer to measure the iodine uptake rate of the thyroid gland. Find (*a*) the disintegration constant λ and (*b*) the half-life $t_{1/2}$ from these data.

Time (min)	*R* (counts/s)	*Time* (min)	*R* (counts/s)
4	392.2	132	10.9
36	161.4	164	4.56
68	65.5	196	1.86
100	26.8	218	1.00

Solution (*a*) If we take the natural logarithm of each side of Eq. 50-7, we find that

$$\ln R = \ln R_0 - \lambda t. \tag{50-9}$$

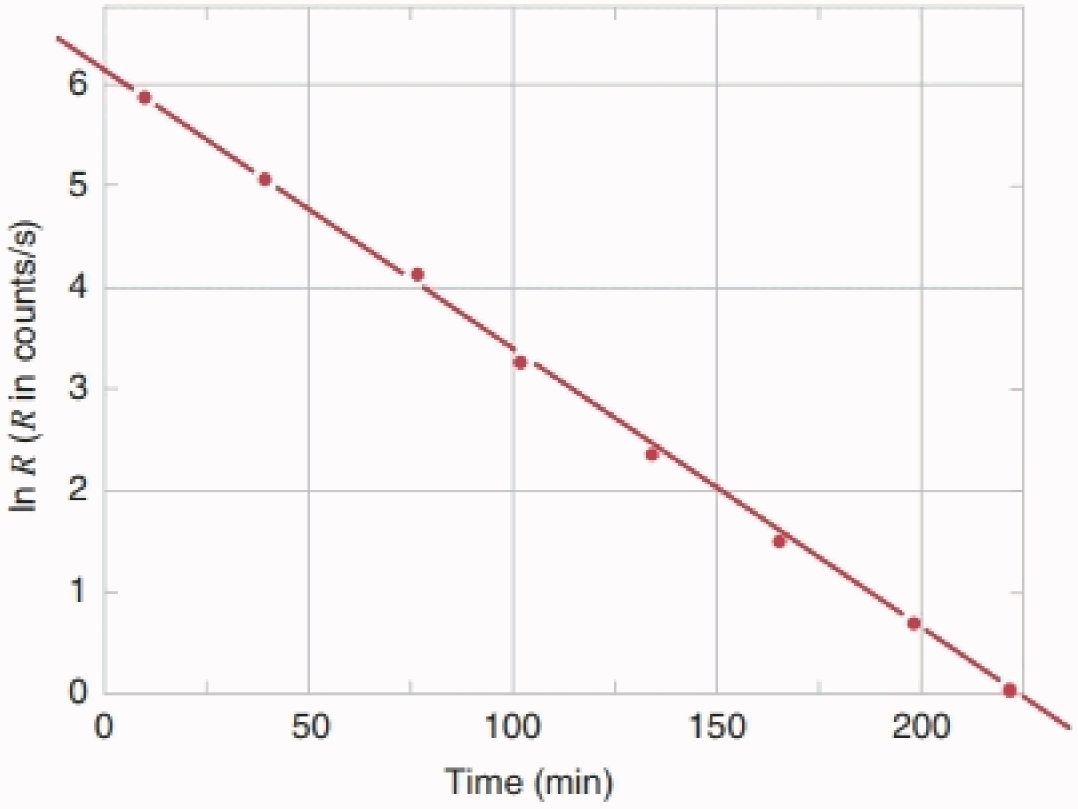


FIGURE 50-8. Sample Problem 50-4. A logarithmic plot of the decay data is fitted by a straight line, showing the exponential nature of the decay. The disintegration constant λ can be found from the slope of the line.

Thus if we plot the natural logarithm of R against t, we should obtain a straight line whose slope is $-\lambda$. Figure 50-8 shows such a plot, using the data given for ^{128}I. Equating the slope of the line to $-\lambda$ yields

$$-\lambda = -\frac{(6.1 - 0)}{(225 \text{ min} - 0)}$$

or

$$\lambda = 0.0271 \text{ min}^{-1}.$$

(*b*) Equation 50-8 yields for $t_{1/2}$:

$$t_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{0.0271 \text{ min}^{-1}} = 25.6 \text{ min}.$$

SAMPLE PROBLEM 50-5. A 1.00-g sample of pure KCl from the chemistry stockroom is found to be radioactive and to decay at an absolute rate R of 1600 counts/s. The decays are traced to the element potassium and in particular to the isotope ^{40}K, which constitutes 1.18% of normal potassium. What is the half-life for this decay?

Solution In the case of long-lived decays, it is not possible to wait long enough to observe a measurable decrease in the decay rate R with time. We must find λ by measuring both N and $-dN/dt$ in Eq. 50-5. The molar mass of KCl is 74.9 g/mol, so the number of potassium atoms in the sample is

$$N_K = \frac{(6.02 \times 10^{23} \text{ mol}^{-1})(1.00 \text{ g})}{74.9 \text{ g/mol}} = 8.04 \times 10^{21}.$$

The number of ^{40}K atoms is 1.18% of N_K, or

$$N_{40} = (0.0118)(8.04 \times 10^{21}) = 9.49 \times 10^{19}.$$

From Eq. 50-5 we have

$$\lambda = \frac{-dN/dt}{N} = \frac{R}{N_{40}} = \frac{1600 \text{ s}^{-1}}{9.49 \times 10^{19}} = 1.69 \times 10^{-17} \text{ s}^{-1},$$

and the half-life, from Eq. 50-8, is

$$t_{1/2} = \frac{\ln 2}{\lambda} = \left(\frac{0.693}{1.69 \times 10^{-17} \text{ s}^{-1}}\right)\left(\frac{1 \text{ y}}{3.16 \times 10^{7} \text{ s}}\right)$$
$$= 1.30 \times 10^{9} \text{ y}.$$

This is of the order of magnitude of the age of the universe. No wonder we cannot measure the half-life of this nuclide by waiting around for its decay rate to decrease! (Interestingly, the potassium in our own bodies has its normal share of the ^{40}K isotope. We are all slightly radioactive.)

50-4 ALPHA DECAY

In a radioactive decay process, the nucleus of an atom with mass m_i undergoes a transformation usually leading to a different nuclide, with the accompanying emission of one or more additional particles. The total mass of all the final particles is m_f. The decay will be possible only if $m_i > m_f$; put another way, according to conservation of energy the total energy before the decay (the rest energy $m_i c^2$ of the initial atom, which we assume to be at rest in our reference frame) must equal the total energy after the decay (the total rest energy $m_f c^2$ of all the final products plus their total kinetic energy K_f):

$$m_i c^2 = m_f c^2 + K_f. \qquad (50\text{-}10)$$

It is convenient to define the *Q-value for the decay* to be the difference between the initial and final rest energies:

$$Q = m_i c^2 - m_f c^2. \qquad (50\text{-}11)$$

In effect, Q is the energy released in the decay. Comparison with Eq. 50-10 shows that, if the initial decaying nucleus is at rest, the Q-value gives the total kinetic energy of all the decay products:

$$Q = K_f. \qquad (50\text{-}12)$$

The energy Q is shared among the decay products consistent with conservation of momentum.

From Eq. 50-12 we can clearly see that the decay is possible only for $Q > 0$, and thus according to Eq. 50-11 m_i must be greater than m_f, as we have already concluded.

Let us apply these concepts to a specific decay process—namely, to alpha decay in which an alpha (α) particle (^{4}He nucleus) is emitted by a nucleus. We consider as an example the decay of ^{238}U. In emitting an α particle, the ^{238}U nucleus loses two protons (from 92 to 90) and two neutrons (from 146 to 144). According to the periodic table, the element with atomic number 90 is Th, and so the complete decay process is

$$^{238}\text{U} \rightarrow {}^{234}\text{Th} + {}^{4}\text{He}. \qquad (50\text{-}13)$$

In Sample Problem 50-6 we show that the decay energy (Q value) for this process is 4.27 MeV, which is shared between the α particle and the recoiling ^{234}Th nucleus. The measured half-life for this decay is 4.47×10^9 y.

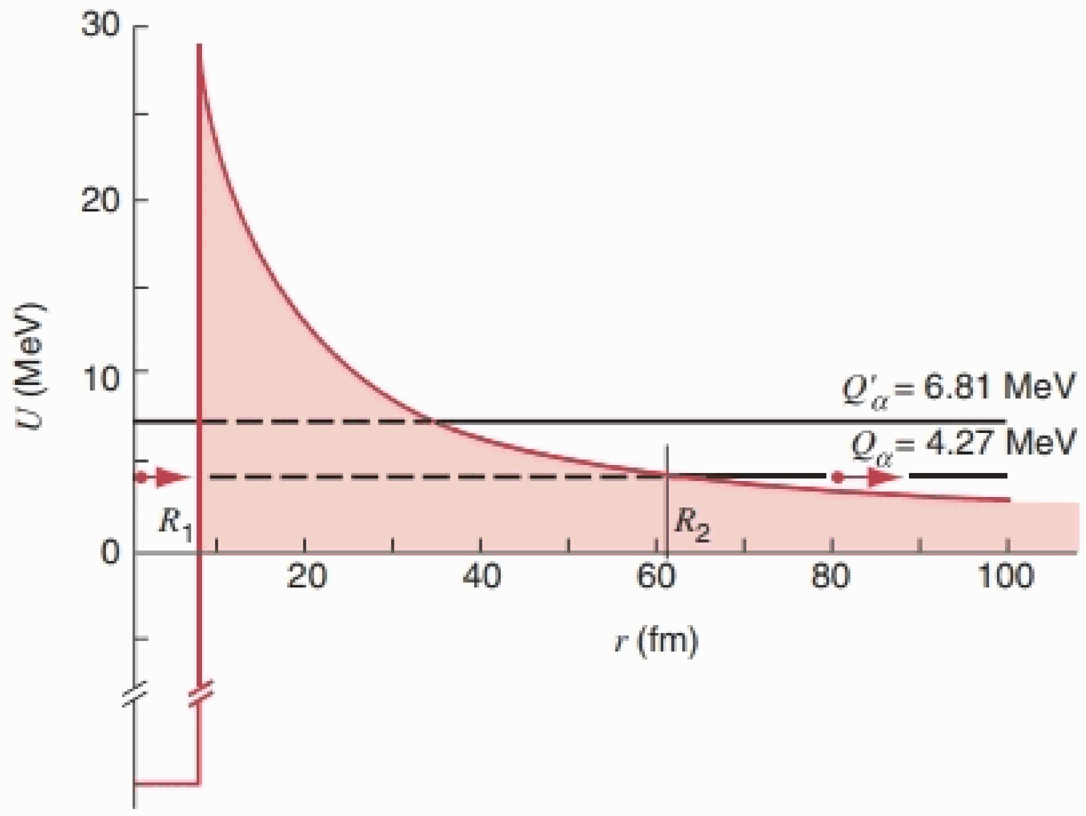


FIGURE 50-9. A potential energy function representing the emission of α particles by ^{238}U. The shaded area represents the potential barrier that inhibits the decay process. The horizontal lines represent the decay energies of ^{238}U (4.27 MeV) and ^{228}U (6.81 MeV).

We now ask ourselves: "If energy is released in every such decay event, why did the ^{238}U nuclei not decay shortly after they were created?" The creation process is believed to have occurred in the violent explosions of ancestral stars (supernovas), predating the formation of our solar system. Why did these nuclei wait so very long before getting rid of their excess energy by emitting an α particle? To answer this question, we must study the detailed mechanism of alpha decay.

We choose a model in which the α particle is imagined to exist preformed inside the nucleus before it escapes. Figure 50-9 shows the approximate potential energy function $U(r)$ for the α particle and the residual ^{234}Th nucleus as a function of their separation. It is a combination of a potential well associated with the (attractive) strong nuclear force that acts in the nuclear interior ($r < R_1$) and a Coulomb potential associated with the (repulsive) electrostatic force that acts between the two particles after the decay has occurred ($r > R_1$).

The horizontal line marked $Q_\alpha = 4.27$ MeV shows the disintegration energy for the process, as calculated in Sample Problem 50-6. Note that this line intersects the potential energy curve at two points, R_1 and R_2. We now see why the α particle is not immediately emitted from the ^{238}U nucleus! That nucleus is surrounded by an impressive potential barrier, shown by the shaded area in Fig. 50-9. Visualize this barrier as a spherical shell whose inner radius is R_1 and whose outer radius is R_2, its volume being forbidden to the α particle under the laws of classical physics. If the α particle found itself in that region, its potential energy U would exceed its total energy E, which would mean, classically, that its kinetic energy K ($= E - U$) would be negative, an impossible situation.

TABLE 50-2 **The Alpha Decay of ^{238}U and ^{228}U**

Nuclide	Q_α	*Half-life*
^{238}U	4.27 MeV	4.5×10^9 y
^{228}U	6.81 MeV	550 s

Indeed, we now change our question and ask: "How can the ^{238}U nucleus *ever* emit an α particle?" The α particle seems permanently trapped inside the nucleus by the barrier.

The answer is that, as we learned in Section 46-7, in wave mechanics there is always a chance (described by Eq. 46-24) that a particle can tunnel through a barrier that is classically insurmountable. In fact, the explanation of alpha decay by wave mechanical barrier tunneling was one of the very first applications of the new quantum physics.

For the long-lived decay of ^{238}U the barrier is actually not very "leaky." We can show that the α particle, presumed to be rattling back and forth within the nucleus, must present itself at the inner surface of the barrier about 10^{38} times before it succeeds in tunneling through. This is about 10^{20} times per second for about 10^9 years! We, of course, are waiting on the outside, taking note of only those α particles that *do* manage to escape.

We can test this barrier-tunneling explanation of alpha decay by looking at other alpha emitters, for which the barrier would be different. For an extreme contrast, consider the alpha decay of another uranium nuclide, ^{228}U, which has a disintegration energy Q'_α of 6.81 MeV, as shown in Fig. 50-9. The barrier in this case is both thinner (compare the lengths of the dashed lines in Fig. 50-9) and lower (compare the heights of the barrier above the dashed lines); if our barrier tunneling notions are correct, we would expect alpha decay to occur more readily for ^{228}U than for ^{238}U. Indeed it does. As Table 50-2 shows, the half-life of ^{228}U is only 550 s! We recall from Section 46-7 that the transmission coefficient of a barrier—because of the exponential nature of Eq. 46-24—is very sensitive to small changes in the dimensions of the barrier. We see that an increase in Q_α by a factor of only 1.6 produces a decrease in half-life (that is, in the effectiveness of barrier tunneling) by a factor of 3×10^{14}.

SAMPLE PROBLEM 50-6. (*a*) Find the energy released during the alpha decay of ^{238}U. (*b*) Show that this nuclide cannot spontaneously emit a proton. The needed atomic masses are

^{238}U	238.050783 u	^{4}He	4.002603 u
^{234}Th	234.043596 u	^{1}H	1.007825 u
^{237}Pa	237.051144 u.		

Solution (*a*) In the alpha decay process of Eq. 50-13 the total atomic mass m_f of the decay products $^{234}Th + {}^{4}He$ is 238.046199 u. According to Eq. 50-11, the Q-value is

$$\begin{aligned} Q &= (m_i - m_f)c^2 \\ &= (238.050783\ \mathrm{u} - 238.046199\ \mathrm{u})(931.5\ \mathrm{MeV/u}) \\ &= 4.27\ \mathrm{MeV}. \end{aligned}$$

This energy is available to share as kinetic energy between the α particle and the recoiling ^{234}Th atom.

(*b*) If ^{238}U were to emit a proton, the decay process would be

$$^{238}\mathrm{U} \rightarrow {}^{237}\mathrm{Pa} + {}^{1}\mathrm{H}.$$

In this case the Q-value is

$$\begin{aligned} Q &= (m_i - m_f)c^2 \\ &= (238.050783\ \mathrm{u} - 238.058969\ \mathrm{u})(931.5\ \mathrm{MeV/u}) \\ &= -7.62\ \mathrm{MeV}. \end{aligned}$$

The negative Q-value shows that ^{238}U is stable against spontaneous proton emission.

50-5 BETA DECAY

A nucleus that decays spontaneously by emitting an electron (either positive or negative) is said to undergo *beta decay.** Here are two examples:

$$^{32}\mathrm{P} \rightarrow {}^{32}\mathrm{S} + \mathrm{e}^- + \bar{\nu} \qquad (t_{1/2} = 14.3\ \mathrm{d}) \qquad (50\text{-}14)$$

and

$$^{64}\mathrm{Cu} \rightarrow {}^{64}\mathrm{Ni} + \mathrm{e}^+ + \nu \qquad (t_{1/2} = 12.7\ \mathrm{h}). \qquad (50\text{-}15)$$

The symbols ν and $\bar{\nu}$ represent the *neutrino* and its antiparticle, the *antineutrino,* neutral particles that are emitted from the nucleus along with the electron or positron (positive electron) during the decay process. Neutrinos interact only very weakly with matter and—for that reason—are so extremely difficult to detect that, for many years, their presence went unnoticed. We consider the fundamental nature and importance of these elusive particles in Chapter 52.

It may seem surprising that nuclei can emit electrons (and neutrinos) in view of the fact that we have said that nuclei are made up of neutrons and protons only. However, we saw earlier that atoms emit photons, and we certainly do not say that atoms "contain" photons. We say that the photons are created during the emission process.

So it is with the electrons and the neutrinos emitted from nuclei during beta decay. They are both created during the emission process, a neutron transforming itself into a proton within the nucleus (or conversely) according to

$$\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e}^- + \bar{\nu} \qquad (\beta^- \text{ decay}) \qquad (50\text{-}16)$$

or

$$\mathrm{p} \rightarrow \mathrm{n} + \mathrm{e}^+ + \nu \qquad (\beta^+ \text{ decay}). \qquad (50\text{-}17)$$

These are the basic beta-decay processes.

In any decay process, the amount of energy released is uniquely determined by the difference in rest energy between the initial nucleus and the final nucleus plus decay products (see Eq. 50-11). In a particular alpha-decay process, such as that of ^{238}U, every emitted α particle carries the same kinetic energy. In beta decay, however, the kinetic energy of the emitted electrons is not uniquely determined. Instead, the emitted electrons have a continuous spectrum of energies, from zero up to a maximum $K_{\max}$, as Fig. 50-10 illustrates for the beta decay of ^{64}Cu (Eq. 50-15).

For many years, before the neutrino was identified, curves such as that of Fig. 50-10 were a challenging puzzle. They suggested that some energy was "missing" in the decay process and led many reputable physicists, including Niels Bohr, to speculate that perhaps the law of conservation of energy might be valid only statistically in such decays.

The answer to this puzzle lies in the emission of the neutrino or antineutrino, which carries a share of the decay energy. If we were to measure the energies of both particles (electron and antineutrino or positron and neutrino) in a particular decay process and add them up, we would come out every time with the same fixed value, equal to the disintegration energy. Energy is indeed conserved in each individual decay process.

The existence of an undetected particle as a solution to the missing energy problem was proposed by Pauli in 1931, and the neutrino was made a part of a formal theory of beta decay by Fermi in 1934. Nevertheless, it took another 20 years before neutrinos were detected in the laboratory. The difficulty in their measurement results from their exceedingly weak interactions with matter—their mean free path through solid matter is of the order of several thousand light years. Today neutrino physics is an important subfield of nuclear and particle physics, and its practitioners study neutrinos emitted from radioactive sources or produced in nuclear reactions on Earth, in the Sun, or in supernovas.

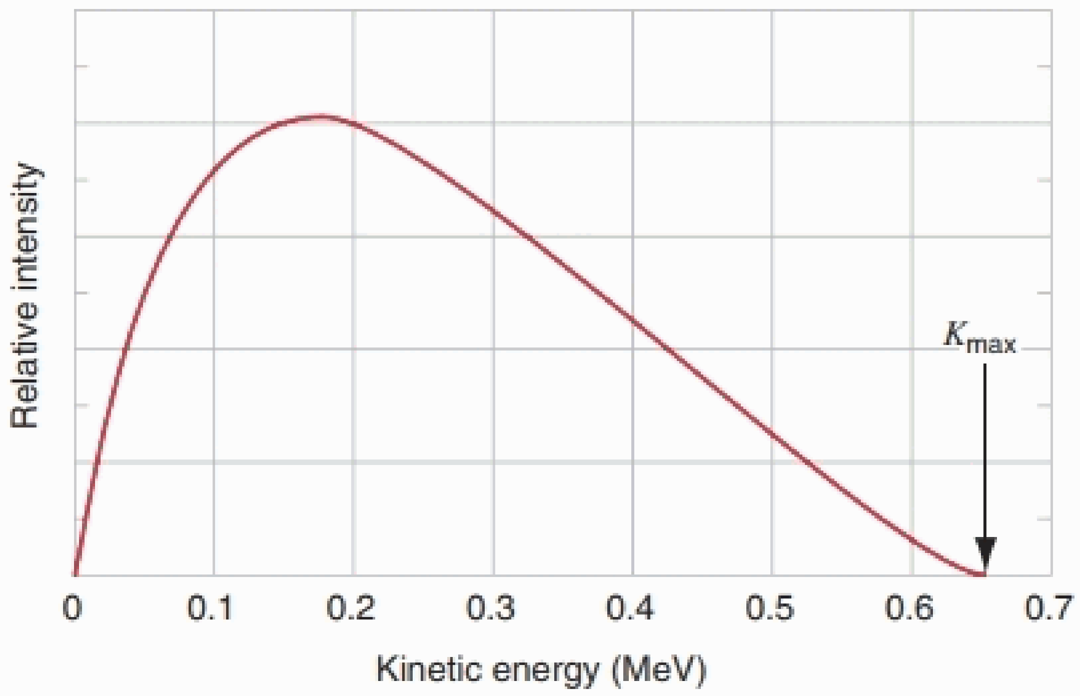


FIGURE 50-10. The kinetic energy distribution of the positrons emitted in the beta decay of ^{64}Cu. The maximum kinetic energy is 0.653 MeV.

---

*Beta decay also includes electron capture, in which a nucleus decays by absorbing one of its orbital electrons. We do not consider that process here.

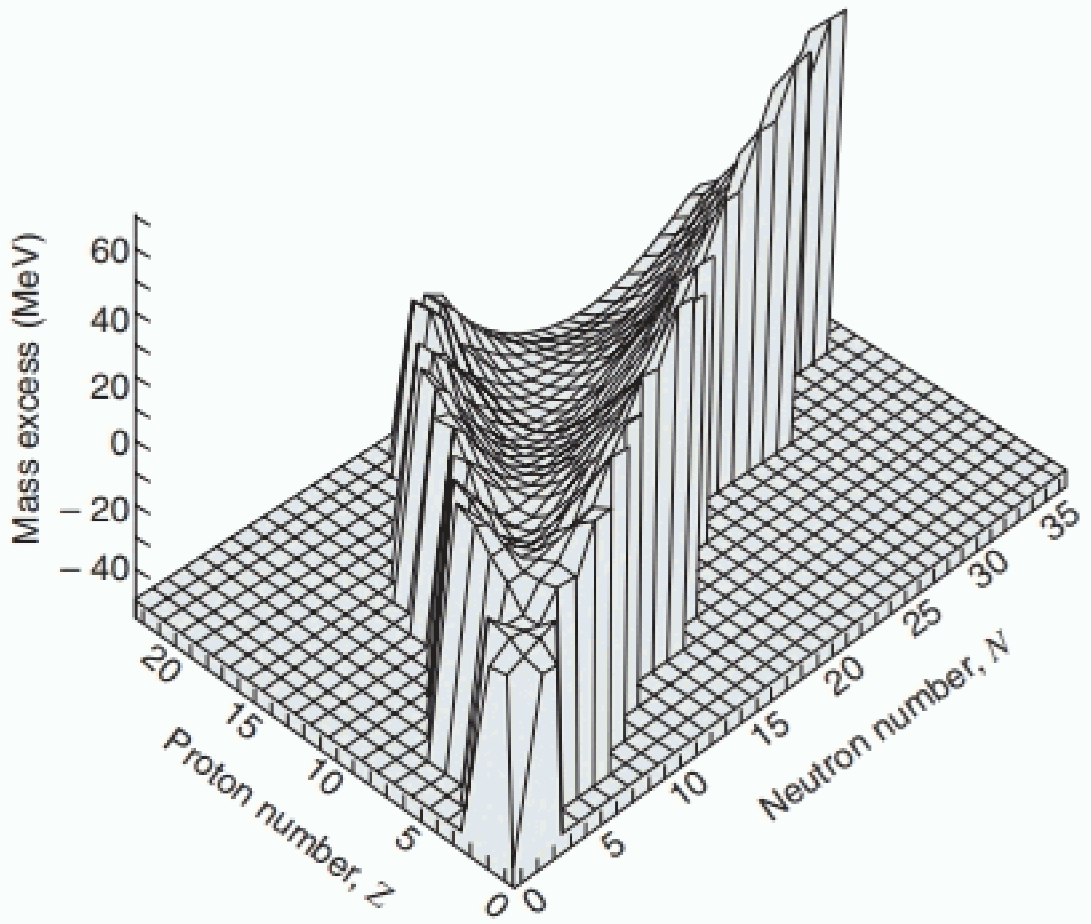


FIGURE 50-11. A portion of the valley of the nuclides, showing only the lightest nuclides. The quantity plotted on the vertical axis is the mass excess, defined as $(m - A)c^2$, where m is the atomic mass in u.

Our study of alpha and beta decay permits us to look at the nuclidic chart of Fig. 50-4 in a new way. Let us construct a three-dimensional surface by plotting the mass of each nuclide in a direction at right angles to the *NZ* plane of that figure. The surface so formed gives a graphic representation of nuclear stability. As Fig. 50-11 shows (for the light nuclides), it describes a "valley of the nuclides," the stability zone of Fig. 50-4 running along its bottom. Nuclides on the headwall of the valley (a region not displayed in Fig. 50-11) decay into it largely by chains of alpha decay and by spontaneous fission. Nuclides on the proton-rich side of the valley decay into it by emitting positive electrons and those on the neutron-rich side do so by emitting negative electrons.

SAMPLE PROBLEM 50-7. Calculate the disintegration energy (*Q*-value) in the beta decay of ^{32}P, as described by Eq. 50-14. The needed atomic masses are 31.973907 u for ^{32}P and 31.972071 u for ^{32}S.

Solution Because of the presence of the emitted electron, we must be especially careful to distinguish between nuclear and atomic masses. Let m' represent the nuclear masses of ^{32}P and ^{32}S, and let m represent their atomic masses. We take the disintegration energy Q to be $\Delta m\, c^2$, where

$$\Delta m = m'(^{32}\text{P}) - [m'(^{32}\text{S}) + m_e],$$

with m_e the mass of the electron and the neutrino assumed to be massless. If we add and subtract $15m_e$ on the right-hand side, we have

$$\Delta m = [m'(^{32}\text{P}) + 15m_e] - [m'(^{32}\text{S}) + 16m_e].$$

The quantities in brackets are the atomic masses. Thus we have

$$\Delta m = m(^{32}\text{P}) - m(^{32}\text{S}).$$

If we subtract the atomic masses in this way, the mass of the emitted electron is automatically taken into account.*

The disintegration energy for the ^{32}P decay is then

$$\begin{aligned} Q &= \Delta m\, c^2 = (31.973907\ \text{u} - 31.972071\ \text{u})(931.5\ \text{MeV/u}) \\ &= 1.71\ \text{MeV}. \end{aligned}$$

This is just equal to the measured value of $K_{\max}$, the maximum energy of the emitted electrons. Thus although 1.71 MeV is released every time a ^{32}P nucleus decays, in essentially every case the electron carries away less energy than this. The neutrino gets the rest, carrying it away from the laboratory undetected. (A negligible share, of the order of eV, also goes to the ^{32}S nucleus in order to conserve momentum in the decay.)

50-6 MEASURING IONIZING RADIATION†

When radiations such as x rays, gamma rays, beta particles, or alpha particles encounter an atom, they can cause the atom to eject electrons and to become ionized. Because ionization can damage individual cells of living tissue, the effects of ionizing radiations have become a matter of general public interest. Such radiations arise in nature from the cosmic rays and also from radioactive elements in the Earth's crust. Artificially produced radiations also contribute, including diagnostic and therapeutic x rays and radiations from radionuclides used in medicine and in industry. The disposal of radioactive waste and the evaluation of the probabilities of nuclear accidents continue to be dealt with at the level of national policy.

It is not our task here to explore the various sources of ionizing radiations but simply to describe the units in which the properties and effects of these radiations are expressed. There are four such units, and they are often used loosely or incorrectly in popular reporting.

1. *The curie (abbreviation Ci).* This is a measure of the *activity* or rate of decay of a radioactive source. It was originally defined as the activity of 1 g of radium in equilibrium with its by-products, but it is now defined simply as

$$1\ \text{curie} = 3.7 \times 10^{10}\ \text{disintegrations per second.}$$

This definition says nothing about the nature of the decays or the amount of energy released in the decay. Note also that this unit is not appropriate to describe the ionizing effects of x rays from, say, a medical x-ray machine. The radiations must be emitted from a radionuclide.

---

*This is not the case for positron decay or for electron capture; see Problems 11 and 12. Note also that in this sample problem we neglect the (small) difference in the binding energies of the atomic electrons before and after the beta decay.

†See "Radiation Exposure in Our Daily Lives," by Stewart C. Bushong, *The Physics Teacher,* March 1977, p. 135.

An example of the proper use of the curie is the statement: "One milligram of ^{239}Pu has an activity of 62 μCi." The fact that ^{239}Pu is an alpha emitter does not enter.

2. *The roentgen (abbreviation R).* This is a measure of *exposure*—that is, of the ability of a beam of x rays or gamma rays to produce ions in a particular substance. Specifically, one roentgen is defined as that exposure that would produce 1.6×10^{12} ion pairs per gram of air, the air being dry and at standard temperature and pressure. We might say, for example: "In 0.1 s, this dental x-ray beam provides an exposure of 30 mR." This says nothing about whether ions are actually produced or whether or not there is a patient in the chair.

3. *The rad.* This is an acronym for *r*adiation *a*bsorbed *d*ose and is a measure, as its name suggests, of the dose actually delivered to a specific object, in terms of the energy transferred to it. An object, which might be a person (whole body) or a specific part of the body (the hands, say) is said to have received an absorbed dose of 1 rad when 10^{-2} J/kg have been delivered to it by ionizing radiations. A typical statement to show the usage is: "A whole-body gamma-ray dose of 300 rad will cause death in 50% of the population exposed to it." By way of comfort we note that the present average exposure to radiation from both natural and artificial sources is about 0.2 rad (= 200 mrad) per year. The SI unit for absorbed dose is the *gray* (Gy), which is equal to 100 rad.

4. *The rem.* This is an acronym for *r*oentgen *e*quivalent in *m*an and is a measure of *dose equivalent.* It takes account of the fact that, although different types of radiation (gamma rays and neutrons, say) may deliver the same energy per unit mass to the body, they do not have the same biological effect. The dose equivalent (in rems) is found by multiplying the absorbed dose (in rads) by a *quality factor* QF, which may be found tabulated in various reference sources. For x rays and electrons, QF = 1. For slow neutrons, QF = 5, and so on. Personnel monitoring devices such as film badges or dosimeters are designed to register the dose equivalent in rems.

An example of correct usage of the rem is: "The recommendation of the National Council on Radiation Protection is that no individual who is (nonoccupationally) exposed to radiations should receive a dose equivalent greater than 500 mrem (=0.5 rem) in any one year." This includes radiations of all kinds, using the appropriate quality factors. The SI unit for dose equivalent is the *sievert* (Sv), where 1 Sv = 100 rem.

SAMPLE PROBLEM 50-8. A dose of 300 rad is lethal to 50% of the population that receives it. If the equivalent amount of energy were absorbed directly as heat, what temperature increase would result? Assume that c, the specific heat capacity of the human body, is the same as that of water—namely, 4180 J/kg · K.

Solution An absorbed dose of 300 rad corresponds to an absorbed energy per unit mass of

$$(300 \text{ rad})\left(\frac{10^{-2} \text{ J/kg}}{1 \text{ rad}}\right) = 3 \text{ J/kg}.$$

The temperature increase that would result from such an influx of heat is found from

$$\Delta T = \frac{Q/m}{c} = \frac{3 \text{ J/kg}}{4180 \text{ J/kg} \cdot \text{K}} = 7.2 \times 10^{-4} \text{ K}.$$

We see from this tiny temperature increase that the damage done by ionizing radiation has very little to do with thermal heating. The harmful effects arise because the ionizing radiation succeeds in breaking molecular bonds and thus interferes with the normal functioning of the tissue in which it has been absorbed.

50-7 NATURAL RADIOACTIVITY

All the elements beyond hydrogen and helium were made in nuclear reactions in the interiors of stars or in explosive supernovas. Both radioactive and stable nuclides are created in these processes. The solar system is composed of nuclides that were formed about 4.5×10^9 years ago. (How this is determined is discussed later in this section.) Most of the radioactive nuclides that were formed at that time have half-lives that are far shorter than a billion years, and so they have long since decayed to stable nuclides through alpha or beta emission. A few of the original radioactive nuclides, however, have half-lives that are not short in comparison with the age of the solar system. The decay of these nuclides can still be observed, and these decays form part of the background of natural radioactivity in our environment.

Some of these radioactive species are part of decay chains that start with heavy nuclides, such as ^{232}Th ($t_{1/2} = 1.4 \times 10^{10}$ y) or ^{238}U ($t_{1/2} = 4.5 \times 10^9$ y). These nuclides decay through a sequence of alpha and beta decays, eventually reaching stable end products (respectively, ^{208}Pb and ^{206}Pb). The intermediate nuclei in these decay chains have much shorter half-lives; the rate at which the original nuclide disappears and is replaced with the stable end product is determined by the longest-lived member of the chain. These decay processes have presumably been going on since the solar system was formed, and so (as we discuss later) the relative amounts of the initial nuclide and stable decay products present in a material can give a measure of the age of the material. These decays are also thought to contribute to the internal heating of the planets.

Normally, the products of these decays remain in place in the rocks or minerals containing the parent nuclide. However, one of the intermediate substances produced in these decay chains, radon, is a gas. Natural decays that occur near the surface of the Earth (and in building materials, such as concrete) release radioactive radon gas into the atmosphere. The hazards of breathing this radon gas are currently the subject of active research. Radon gas can also be released from the fracture of rocks beneath the surface; therefore the detection of radon gas has been explored as a way of predicting earthquakes.

TABLE 50-3 **Some Natural Radioactive Isotopes**

Isotope	$t_{1/2}$ (y)
^{40}K	1.28×10^9
^{87}Rb	4.8×10^{10}
^{113}Cd	9×10^{15}
^{115}In	4.4×10^{14}
^{138}La	1.3×10^{11}
^{176}Lu	3.6×10^{10}
^{187}Re	5×10^{10}

In addition to the heavy elements, other long-lived radioactive nuclides are present in natural substances. Some of these are listed in Table 50-3.

Other radioactive nuclides are continually produced by natural processes, generally in the Earth's atmosphere by reactions of molecules of the air with cosmic rays (high-energy protons from space). Notable among these is ^{14}C ($t_{1/2} = 5730$ y), which has important applications in radioactive dating of organic materials.

Radioactive Dating

Suppose we have an initial radionuclide I that decays to a final product F with a known half-life $t_{1/2}$. At a particular time $t = 0$, we start with N_0 initial nuclei and none of the final product nuclei. At a later time t, we find N_I of the original nuclei remain, while N_F $(= N_0 - N_I)$ of the product nuclei have appeared. The initial nuclei decay according to

$$N_I = N_0 e^{-\lambda t},$$

and thus

$$t = \frac{1}{\lambda} \ln \frac{N_0}{N_I} = \frac{t_{1/2}}{\ln 2} \ln \frac{N_0}{N_I}$$

or, substituting $N_I + N_F$ for N_0,

$$t = \frac{t_{1/2}}{\ln 2} \ln \left(1 + \frac{N_F}{N_I}\right). \tag{50-18}$$

That is, a measurement of the present ratio of product and original nuclei can determine the age of the sample.

This calculation has been based on the assumption that none of the product nuclei were present at $t = 0$. This assumption may not always be valid, but there are techniques for radioactive dating that can correct for the presence of these original product nuclei.

This method can be used to determine the time since the formation of the solar system; examples include the ratios of ^{238}U to ^{206}Pb, ^{87}Rb to ^{87}Sr, and ^{40}K to ^{40}Ar. Terrestrial rocks, Moon rocks, and meteorites analyzed by these methods all seem to have common ages of around 4.5×10^9 y, which we take to be the age of the solar system.

The radioactive isotope ^{14}C is present in the atmosphere; about 1 carbon atom in 10^{12} is radioactive ^{14}C. Each gram of carbon has an activity of about 12 decays per minute due to ^{14}C. Living organisms can absorb this activity by aspiration of CO_2 or by eating plants that have done so. When the organism dies, it stops absorbing ^{14}C, and the ^{14}C present at its death begins to decay. By measuring the decay rate of ^{14}C, we can determine the age of the sample. For example, if we examine a sample and it shows 6 decays per minute per gram of carbon, we know that the original activity has been reduced by half, and the sample must be one half-life (5730 y) old.

This method of *radiocarbon dating* (which was developed in 1947 by Willard Libby, who was awarded the 1960 Nobel Prize in chemistry for this work) is useful for samples of organic matter that are less than about 10 half-lives in age. In 10 half-lives, the activity of a sample drops by a factor of 2^{-10}, or about 10^{-3}, and the decay rate becomes too small to be determined with precision. The practical upper limit on the age of samples that can be dated by this method is about 50,000 y. In recent years, a new technique has been developed in which an accelerator is used as a mass spectrometer to determine the ^{14}C/^{12}C ratio to high precision. In this way the usefulness of radiocarbon dating has been extended to samples as old as 100,000 y.

SAMPLE PROBLEM 50-9. In a sample of rock, the ratio of ^{206}Pb to ^{238}U nuclei is found to be 0.65. What is the age of the rock?

Solution From Eq. 50-18, using 4.5×10^9 y for the half-life of ^{238}U, we have

$$t = \frac{4.5 \times 10^9 \text{ y}}{0.693} \ln (1 + 0.65) = 3.3 \times 10^9 \text{ y}.$$

This rock is somewhat younger than the maximum age of 4.5×10^9 y that we determine for rocks in the solar system, which may suggest that the rock did not solidify until 3.3×10^9 y ago. The ^{206}Pb that was formed prior to that time was "boiled off" from the molten rock. Only after the rock solidified could the ^{206}Pb begin to accumulate.

50-8 NUCLEAR REACTIONS

We can represent a nuclear reaction by

$$X + a \rightarrow Y + b \tag{50-19}$$

or, in more compact notation,

$$X(a,b)Y. \tag{50-20}$$

Usually, particle a is the *projectile nucleus* and particle X is the *target nucleus,* which is often at rest in the laboratory. If the projectile is a charged particle, it may be raised to its desired energy in a Van de Graaff accelerator (see Section 28-10) or a cyclotron (see Section 32-3). The projectile may also be a neutron from a nuclear reactor. It is customary to designate product particle Y as the heavier *residual nucleus* and b as the lighter *emerging nucleus.*

The reaction energy Q is defined in analogy with Eq. 50-11 as $m_i c^2 - m_f c^2$, or

$$Q = (m_X + m_a)c^2 - (m_Y + m_b)c^2. \qquad (50\text{-}21)$$

Using energy conservation, we can write Eq. 50-21 as

$$Q = (K_Y + K_b) - (K_X + K_a), \qquad (50\text{-}22)$$

in which K represents the kinetic energy.

Equations 50-21 and 50-22 are valid only when Y and b are in their ground states. As we discuss later in this section, if either nuclide is produced in an excited state, the reaction energy is reduced by the excitation energy.

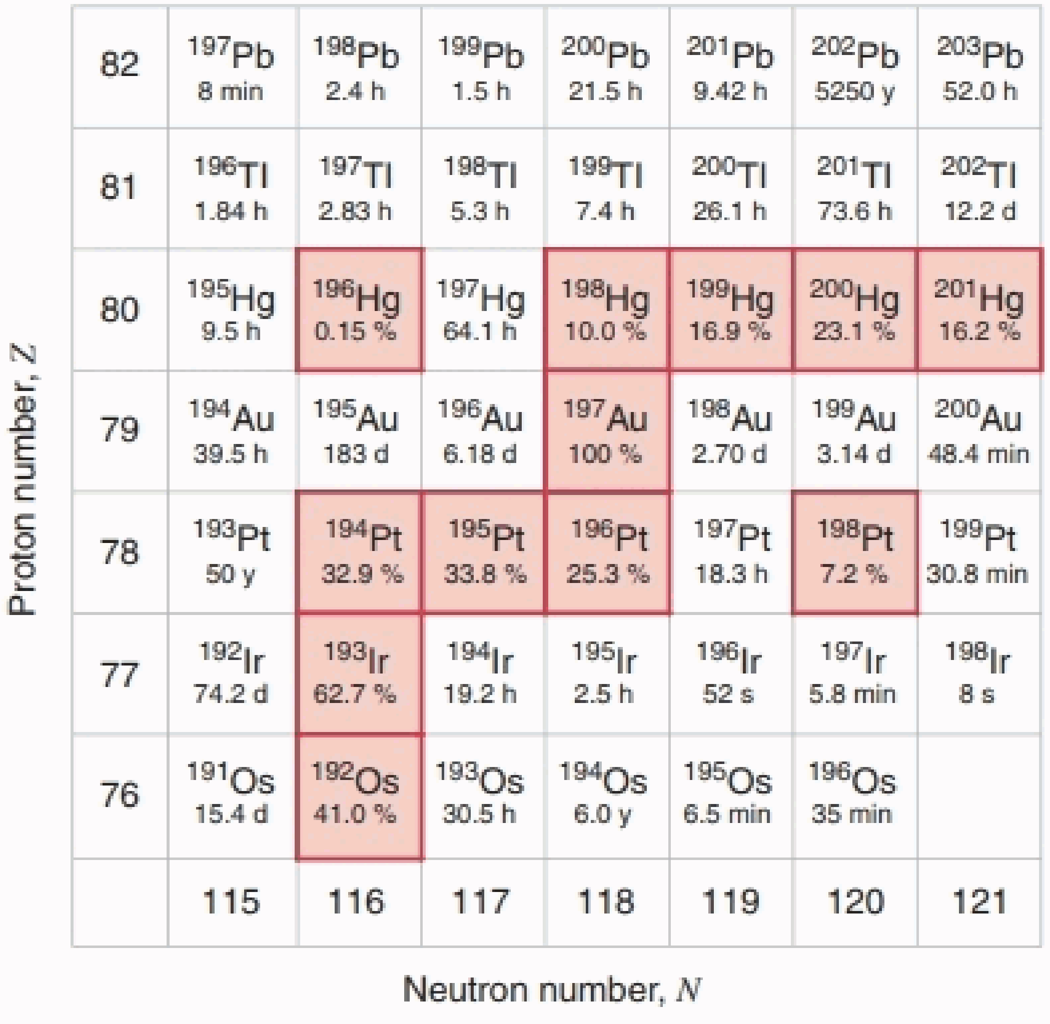
A typical reaction is

$$^{19}\mathrm{F}(\mathrm{p},\alpha)^{16}\mathrm{O},$$

for which $Q = 8.13$ MeV. Equations 50-21 and 50-22 tell us that, in this reaction, the system loses rest energy and gains kinetic energy, in amount 8.13 MeV per event. Reactions, like this one, for which $Q > 0$ are called *exothermic*. Reactions for which $Q < 0$ are called *endothermic;* such reactions will not "go" unless a certain minimum kinetic energy (the *threshold energy*) is carried into the system by the projectile.

If a and b are identical particles, which requires that X and Y also be identical, we describe the reaction as *scattering*. If the kinetic energy of the system is the same both before and after the event (which means that $Q = 0$ and all nuclides remain in their ground states), we have *elastic scattering*. If these energies are different ($Q \neq 0$), we have *inelastic scattering,* in which case Y or b may be left in an excited state.

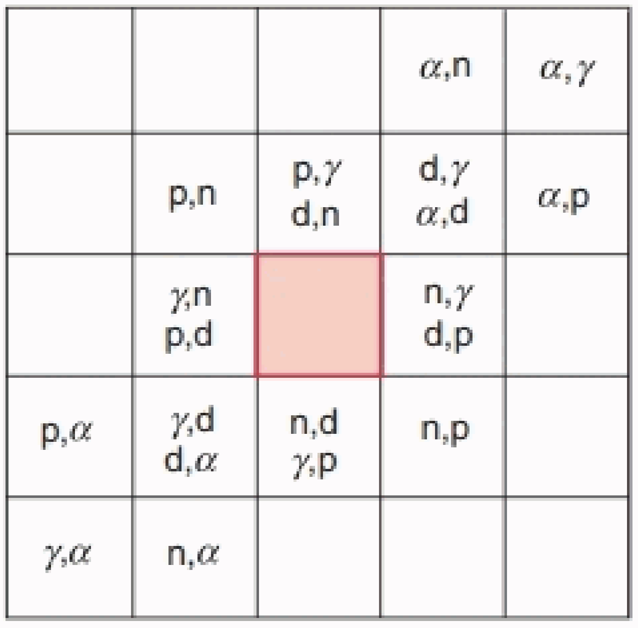
We can easily keep track of nuclear reactions by plotting them on a nuclidic chart like that of Fig. 50-4. Figure 50-12 shows an enlarged portion of such a chart, centered arbitrarily on the nuclide ^{197}Au. Stable nuclides are shaded, and their isotopic abundances are shown. The unshaded squares represent radionuclides, with their half-lives shown.



Proton number, Z							
82	^{197}Pb 8 min	^{198}Pb 2.4 h	^{199}Pb 1.5 h	^{200}Pb 21.5 h	^{201}Pb 9.42 h	^{202}Pb 5250 y	^{203}Pb 52.0 h
81	^{196}Tl 1.84 h	^{197}Tl 2.83 h	^{198}Tl 5.3 h	^{199}Tl 7.4 h	^{200}Tl 26.1 h	^{201}Tl 73.6 h	^{202}Tl 12.2 d
80	^{195}Hg 9.5 h	^{196}Hg 0.15 %	^{197}Hg 64.1 h	^{198}Hg 10.0 %	^{199}Hg 16.9 %	^{200}Hg 23.1 %	^{201}Hg 16.2 %
79	^{194}Au 39.5 h	^{195}Au 183 d	^{196}Au 6.18 d	^{197}Au 100 %	^{198}Au 2.70 d	^{199}Au 3.14 d	^{200}Au 48.4 min
78	^{193}Pt 50 y	^{194}Pt 32.9 %	^{195}Pt 33.8 %	^{196}Pt 25.3 %	^{197}Pt 18.3 h	^{198}Pt 7.2 %	^{199}Pt 30.8 min
77	^{192}Ir 74.2 d	^{193}Ir 62.7 %	^{194}Ir 19.2 h	^{195}Ir 2.5 h	^{196}Ir 52 s	^{197}Ir 5.8 min	^{198}Ir 8 s
76	^{191}Os 15.4 d	^{192}Os 41.0 %	^{193}Os 30.5 h	^{194}Os 6.0 y	^{195}Os 6.5 min	^{196}Os 35 min	
	115	116	117	118	119	120	121

Neutron number, N

FIGURE 50-12. An expanded portion of the chart of the nuclides (Fig. 50-4).



			α,n	α,γ
	p,n	p,γ d,n	d,γ α,d	α,p
	γ,n p,d	▨	n,γ d,p	
p,α	γ,d d,α	n,d γ,p	n,p	
γ,α	n,α			

FIGURE 50-13. Placing this as an overlay on Fig. 50-12, with the shaded central square over a particular target nuclide, shows the residual nuclides that result from the indicated reactions.

Figure 50-13 suggests a transparent overlay that we can place over a nuclidic chart such as that of Fig. 50-12. If the shaded central square of Fig. 50-13 overlays a particular target on the chart of Fig. 50-12, the residual nuclides resulting from the various reactions printed on the overlay are identified.

Thus if we chose ^{197}Au as a target, a (p,α) reaction will produce (stable) ^{194}Pt, and either an (n,γ) or a (d,p) reaction will produce the radionuclide ^{198}Au, whose half-life is 2.70 d.

Nuclei, like atoms, have stationary states of definite energy, and reaction studies can be used to identify them. Consider, for example, the reaction

$$^{27}\mathrm{Al}(\mathrm{d},\mathrm{p})^{28}\mathrm{Al}, \qquad Q = 5.49\ \mathrm{MeV},$$

in which a thin aluminum target foil is bombarded with 2.10-MeV deuterons. In the laboratory the emerging protons are seen to come off with a number of well-defined discrete energies and are accompanied by gamma rays. Fig-

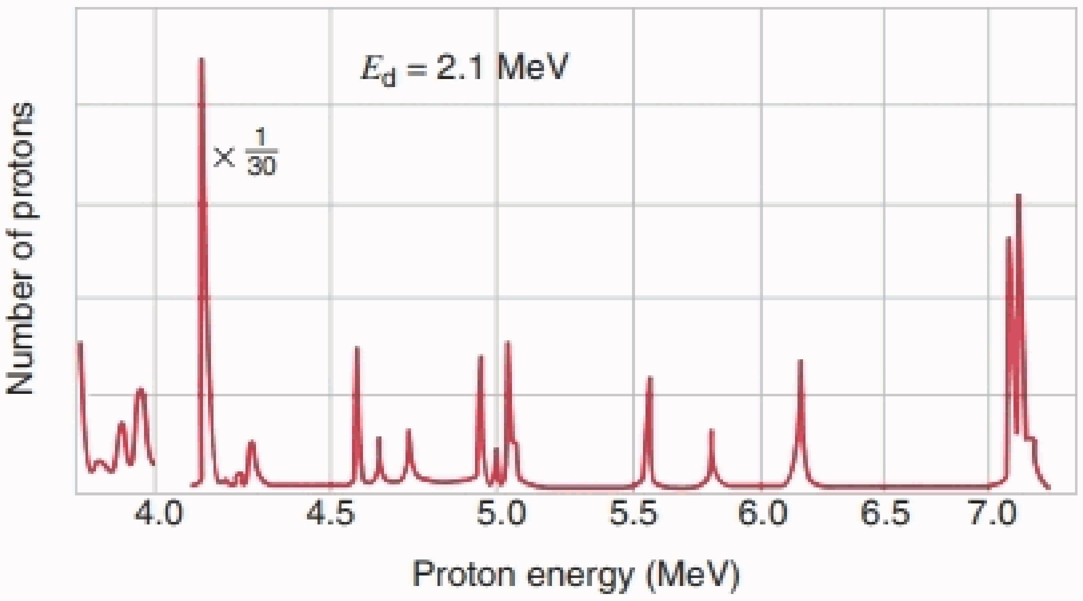


FIGURE 50-14. The energy distribution of protons resulting from the reaction ^{27}Al(d,p)^{28}Al. The incident deuteron has an energy of 2.10 MeV. The protons are detected as they emerge from the target at right angles to the incident beam.

ure 50-14 shows the energy distribution of the emerging protons.

In every reaction event we know that an energy equal to the kinetic energy of the incident deuteron (= 2.10 MeV) plus the reaction energy Q (= 5.49 MeV) is available to be shared between the two reaction products—that is, between the residual nucleus ^{28}Al and the emerging proton p. How is this total energy (2.10 MeV + 5.49 MeV = 7.59 MeV) to be shared between these two particles?

It all depends on whether the residual nucleus ^{28}Al is produced in its ground state or in one of its excited stationary states. In the former case, the emerging proton will have the maximum possible energy, corresponding to the peak on the extreme right of the proton spectrum in Fig. 50-14. If, however, the residual nucleus is formed in an excited state, that nucleus will retain more of the available energy and there will be less energy left for the emerging proton. The residual nucleus will not remain in its excited state very long but will get rid of its excess energy, such as by emitting a gamma ray.

Every proton peak in the spectrum of Fig. 50-14 corresponds to a stationary state of the residual nucleus ^{28}Al. Figure 50-15 shows the energy levels that may be deduced by analyzing this spectrum. You can see the correspondence between the peaks of Fig. 50-14 and the energy levels of Fig. 50-15. We have seen that our understanding of the way atoms are put together rests on the measured energies of the hydrogen atom states as its firm foundation. In the same way, we can learn how nuclei are put together by studying the energies and other properties of their stationary states.

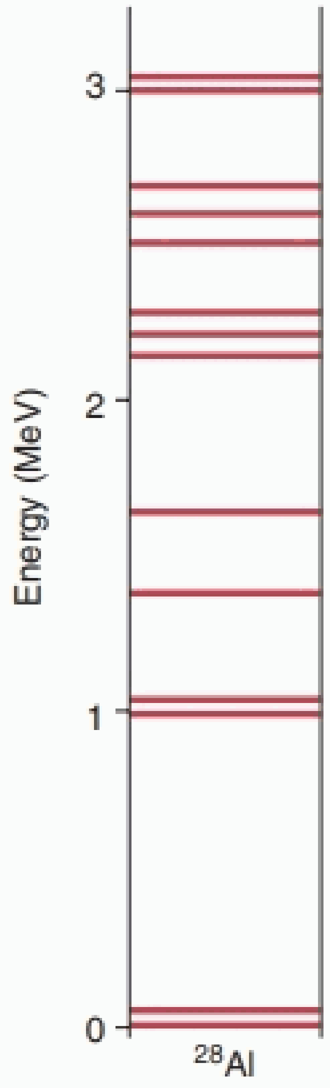


FIGURE 50-15. Energy levels of ^{28}Al, deduced from data such as those of Fig. 50-14.

SAMPLE PROBLEM 50-10. In the reaction

$$^1\text{H} + {}^3\text{H} \rightarrow {}^2\text{H} + {}^2\text{H},$$

protons (^{1}H) with kinetic energy 5.70 MeV are incident on ^{3}H at rest. (*a*) What is the Q value for this reaction? (*b*) Find the kinetic energies of the deuterons emitted along the direction of the incident proton.

Solution (*a*) From Eq. 50-21 we have

$$\begin{aligned} Q &= [m(^1\text{H}) + m(^3\text{H}) - m(^2\text{H}) - m(^2\text{H})]c^2 \\ &= (1.007825\ \text{u} + 3.016049\ \text{u} - 2.014102\ \text{u} \\ &\quad - 2.014102\ \text{u})(931.5\ \text{MeV/u}) \\ &= -4.03\ \text{MeV}. \end{aligned}$$

This reaction is endothermic; the final products have the greater mass and correspondingly the smaller kinetic energy by Eq. 50-22.

(*b*) Using Eq. 50-22, with $K = 0$ for the initial ^{3}H, we have

$$\begin{aligned} K_1 + K_2 = Q + K_p &= -4.03\ \text{MeV} + 5.70\ \text{MeV} \\ &= 1.67\ \text{MeV}. \end{aligned} \qquad (50\text{-}23)$$

Here the subscripts 1 and 2 refer to the two ^{2}H product nuclei. Conservation of momentum along the direction of the incident protons gives

$$\begin{aligned} p_1 + p_2 = p_p &= \sqrt{2m(^1\text{H})K_p} = \sqrt{2(938\ \text{MeV}/c^2)(5.70\ \text{MeV})} \\ &= 103.4\ \text{MeV}/c. \end{aligned} \qquad (50\text{-}24)$$

Equations 50-23 and 50-24 can be solved as two equations in two unknowns (either p_1 and p_2 or K_1 and K_2). The results are

$$K_1 = 0.24\ \text{MeV} \quad \text{and} \quad K_2 = 1.43\ \text{MeV}.$$

Note that we have used nonrelativistic dynamics in solving this problem. Is this a good approximation?

50-9 NUCLEAR MODELS (Optional)

The structure of atoms is now well understood. The Coulomb force is exerted by the massive center (the nucleus) on the electrons, and (given enough computer time) we can use the methods of quantum mechanics to calculate properties of the atom.

Things are not quite so well understood in the case of nuclei. The force law is complicated and cannot be written down explicitly in full detail. Nor is there a natural force center to simplify the calculations. To understand nuclear structure, we face a many-body problem of great complexity.

In the absence of a comprehensive theory of nuclear structure, we try instead to construct *nuclear models.* Physicists use models as simplified ways of looking at a complex system to give physical insight into its properties. The usefulness of a model is tested by its ability to make predictions that can be verified experimentally in the laboratory.

Two models of the nucleus have proved useful. One model (the *collective model*) describes situations in which we can consider all the protons and neutrons to behave cooperatively. The other model (the *independent particle* model) neglects all but one proton or neutron in determining the properties of the nucleus. These two models represent quite opposing views of nuclear structure, but they can be combined to create a single unified model of the nucleus.

The Collective Model

In the collective model, we ignore the motions of individual nucleons and treat the nucleus as a single entity. This model, originally called the "liquid drop model," was developed by Niels Bohr to explain nuclear fission. We imagine the nucleus as a body analogous to a liquid drop, in which the nucleons interact with each other like molecules in the liquid.

The equilibrium shape of the liquid drop is determined by the interactions of its molecules, and similarly the equilibrium shape of a nucleus is determined by the interactions of its nucleons. Many nuclei have spherical equilibrium shapes, whereas others may be ellipsoidal.

Like a liquid drop, a nucleus can absorb energy by the entire nucleus rotating about an axis or vibrating about its equilibrium shape. Through radioactive decay or nuclear reaction experiments, it is possible to study these excited states. Figure 50-16 shows examples of the two kinds of situations. The rotational energy $\frac{1}{2}I\omega^2$ can be written in terms of the angular momentum $L\ (=I\omega)$ as $L^2/2I$. Writing the quantized angular momentum according to Eq. 47-28 as $L = \sqrt{J(J+1)}(h/2\pi)$, where J is the rotational angular momentum quantum number of the entire nucleus, we obtain

$$E_J = \frac{h^2}{8\pi^2 I} J(J+1). \qquad (50\text{-}25)$$

Note that the spacing between the states grows as the angular momentum increases.

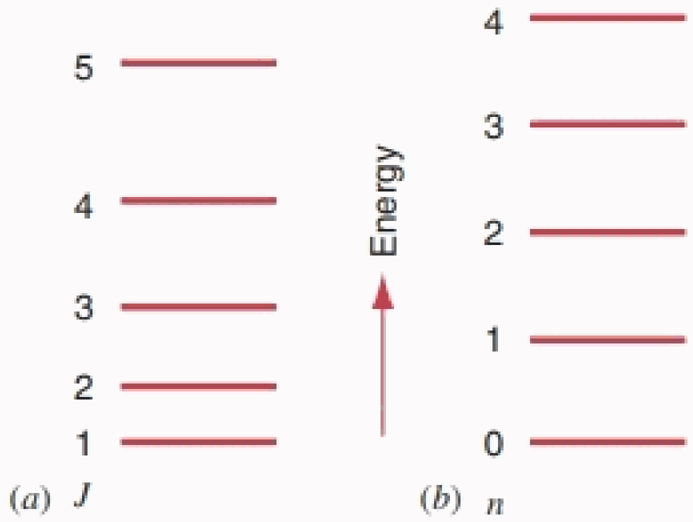


FIGURE 50-16. (*a*) Rotational excited states, labeled with the angular momentum quantum number *J*. (*b*) Vibrational states, labeled with the vibrational quantum number *n*.

The vibrational states have energies given by

$$E_n = nhf \qquad n = 1, 2, 3, \ldots, \qquad (50\text{-}26)$$

where f is the vibrational frequency. Typically the vibrational energy quantum hf has an energy of about 0.5 MeV, corresponding to a frequency of the order of 10^{20} Hz. Note that the vibrational states shown in Fig. 50-16*b* are equally spaced, in contrast to the rotational states.

Experimental evidence for collective structure in nuclei includes not only the observation of rotational and vibrational excited states similar to those of Fig. 50-16, but also nuclear reaction processes in which the energy is shared by all of the nucleons in the nucleus, gamma-ray emission probabilities, and nonspherical (often ellipsoidal) equilibrium shapes of many nuclei.

The Independent-Particle Model

At the other extreme, the independent-particle model assumes that each nucleon can be assigned well-defined states similar to the states of electrons in atoms. Because neutrons and protons, like electrons, must obey the Pauli exclusion principle, the independent-particle structure of a nucleus is very similar to the structure for atoms we discussed in Chapter 48. Like electrons, nucleons arrange themselves into shells with well-defined quantum numbers. When a shell is filled, the result is a nucleus of unusual stability, in analogy with the inert gases that correspond to filled atomic shells.

The closed shells for nucleons occur at so-called "magic" neutron or proton numbers of 2, 8, 20, 28, 50, 82, and 126; these differ slightly from the closed atomic shells (2, 10, 18, 36, 54, and 86) due in part to the differences between the nuclear force and the Coulomb force that acts on the electrons. Some nuclei, called "doubly magic," have closed shells of both protons and neutrons and are therefore especially stable—for example, ^{40}Ca with 20 protons and 20 neutrons. To remove a proton from ^{40}Ca takes 8.3 MeV; however, to remove a proton from a nucleus with one additional proton (^{41}Sc, with a nonmagic 21 protons), takes only 1.1 MeV. Similarly, to remove a neutron from ^{40}Ca takes 15.6 MeV, but only 8.4 MeV is required to remove a neutron from ^{41}Ca, with nonmagic 21 neutrons. This sudden increase in the energy to remove a particle from a closed shell is also observed in atoms (see Fig. 48-6 for the ionization energies of atoms). A similar effect occurs in the ionic radii of atoms, which show a sudden increase as one electron shell is filled and the next electron must start a new shell. Nuclear radii show the same shell effect, as illustrated in Fig. 50-17. Note the sudden jumps in the radius at $N = 20$, 28, 50, 82, and 126, corresponding to filled neutron shells.

As in the case of the nuclear radius or the energy needed to remove one proton or neutron, the independent-particle model at its extreme limit assumes that the properties of the nucleus are determined by a single "valence"

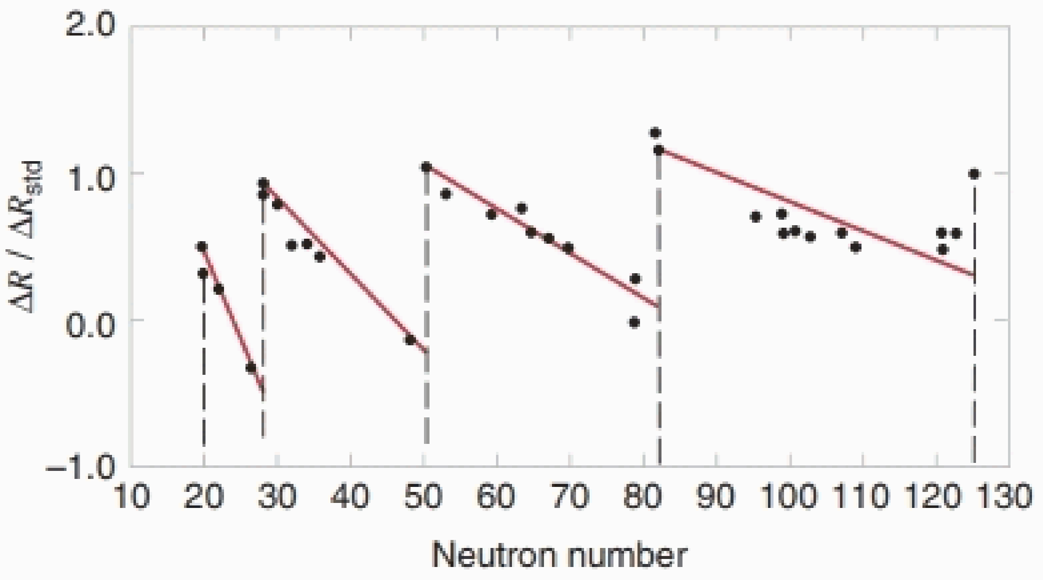


FIGURE 50-17. The variation in nuclear radius as a function of neutron number. The variation is expressed relative to the "standard" variation expected from the "collective" structure of $R = R_0 A^{1/3}$. The sudden jumps indicate shell structure.

proton or neutron. Other nuclear properties that can be successfully interpreted in the independent-particle model include alpha-decay half-lives, magnetic dipole moments, reaction probabilities for capturing a single particle such as a neutron, and energies of excited states.

SAMPLE PROBLEM 50-11. The nuclide ^{120}Sn ($Z = 50$) has a filled proton shell, 50 being one of the magic nucleon numbers. The nuclide ^{121}Sb ($Z = 51$) has an "extra" proton outside this shell. According to the shell concept, this extra proton should be easier to remove than a proton from the filled shell. Verify this by calculating the required energy in each case. Use the following mass data:

Nuclide	*Z*	*N*	*Atomic Mass* (u)
^{121}Sb	50 + 1	70	120.903818
^{120}Sn	50	70	119.902197
^{119}In	50 − 1	70	118.905846

The proton atomic mass is 1.007825 u.

Solution Removing the "extra" proton corresponds to the process

$$^{121}\text{Sb} \rightarrow {}^{120}\text{Sn} + \text{p}.$$

The required energy E follows from

$$\begin{aligned} E &= [m(^{120}\text{Sn}) + m(^{1}\text{H}) - m(^{121}\text{Sb})]c^2 \\ &= (119.902197\ \text{u} + 1.007825\ \text{u} - 120.903818\ \text{u}) \\ &\quad \times (931.5\ \text{MeV/u}) \\ &= 5.8\ \text{MeV}. \end{aligned}$$

Removing the proton from the filled shell corresponds to

$$^{120}\text{Sn} \rightarrow {}^{119}\text{In} + \text{p}.$$

The required energy follows from

$$\begin{aligned} E &= [m(^{119}\text{In}) + m(^{1}\text{H}) - m(^{120}\text{Sn})]c^2 \\ &= (118.905846\ \text{u} + 1.007825\ \text{u} - 119.902197\ \text{u}) \\ &\quad \times (931.5\ \text{MeV/u}) \\ &= 10.7\ \text{MeV}. \end{aligned}$$

This is considerably greater than the energy required to remove an "extra" proton (= 5.8 MeV), just as the shell model predicts. In much the same way, the energy needed to remove an *electron* from a filled *electron shell* (= 22 eV for the filled shell of neon) is much greater than that needed to remove an "extra" electron from outside such a filled shell (= 5 eV for the "extra" electron from sodium).

MULTIPLE CHOICE

50-1 Discovering the Nucleus

1. Rutherford was able to ignore the effect of the electrons in his analysis of alpha-particle scattering experiments because
  (A) electrons do not exert a force on alpha particles.
  (B) the mass of an electron is much less than the mass of an alpha particle.
  (C) the electrons are uniformly distributed throughout the atom.
  (D) the electrons are in such rapid motion that an alpha particle cannot collide with them.

50-2 Some Nuclear Properties

2. How many neutrons are in the nuclide ^{66}Zn?
  (A) 26 (B) 30 (C) 36 (D) 66

3. The binding energy for nucleus A is 7.7 MeV, and that for nucleus B is 7.8 MeV. Which nucleus has the larger mass?
  (A) Nucleus A (B) Nucleus B
  (C) More information is needed to answer.

4. If a nucleus were as big as a grape, an atom would be as big as
  (A) a house. (B) a football field.
  (C) a city. (D) the Moon.

50-3 Radioactive Decay

5. The half-life of a certain radioactive sample is 30 minutes. At 2:00 P.M. the decay rate is measured to be 1200/s. What will be the outcome of a measurement of the decay rate at 3:00 P.M. on that same day?
  (A) 4800/s (B) 1200/s (C) 600/s (D) 300/s

50-4 Alpha Decay

6. A certain heavy nucleus alpha decays with a disintegration energy of 1.50 MeV. From this we can conclude that the kinetic energies observed for the alpha particles
  (A) are all equal to 1.50 MeV.
  (B) are all a little less than 1.50 MeV.
  (C) could occasionally be greater than 1.50 MeV.
  (D) vary continuously from zero to 1.50 MeV.

50-5 Beta Decay

7. A certain stable nuclide, after absorbing a neutron, emits a negative electron and then splits spontaneously into two alpha particles. The original nuclide is
  (A) ^{8}Be. (B) ^{7}Be. (C) ^{7}B. (D) ^{7}Li.

8. In a beta-decay experiment, an electron is observed with a kinetic energy of 1.0 MeV. From this observation, what can be concluded about the disintegration energy Q of the decay?
(A) $Q = 1.0$ MeV
(B) $Q \leq 1.0$ MeV
(C) $Q \geq 1.0$ MeV
(D) Nothing can be concluded about Q.

50-6 Measuring Ionizing Radiation

9. The decay rate of a radioactive source is measured in units of ________, and the biological effect of that radiation on a human being is measured in units of ________.
(A) curies (B) roentgens (C) rads (D) rems

50-7 Natural Radioactivity

10. The decay chain that leads from $^{238}_{92}$U to $^{206}_{82}$Pb consists of a series of alpha decays and beta decays. How many alpha particles are emitted?
(A) 4 (B) 5 (C) 6 (D) 8

50-8 Nuclear Reactions

11. In a nuclear reaction, a beam of alpha particles (^{4}He) strikes a target of ^{60}Ni. The products of this reaction might be
(A) ^{63}Zn + n. (B) ^{63}Cu + p. (C) ^{61}Ni + ^{3}He.
(D) All of the above

50-9 Nuclear Models

12. How many elements are found in nature that have filled electron shells in the atom as well as filled proton and neutron shells in the nucleus?
(A) 1 (B) 2 (C) 3 (D) 4

QUESTIONS

1. When a thin foil is bombarded with α particles, a few of them are scattered back toward the source. Rutherford concluded from this that the positive charge of the atom—and also most of its mass—must be concentrated in a very small "nucleus" within the atom. What was his line of reasoning?

2. In what ways do the so-called strong force and the electrostatic or Coulomb force differ?

3. Why does the *relative* importance of the Coulomb force compared to the strong nuclear force increase at large mass numbers?

4. In your body, are there more neutrons than protons? More protons than electrons? Discuss.

5. Why do nuclei tend to have more neutrons than protons at high mass numbers?

6. Why do we use atomic rather than nuclear masses in analyzing most nuclear decay and reaction processes?

7. How might the equality $1\ \text{u} = 1.6605 \times 10^{-27}$ kg be arrived at in the laboratory?

8. The atoms of a given element may differ in mass, have different physical characteristics, and yet not vary chemically. Why is this?

9. The deviation of isotopic masses from integer values is due to many factors. Name some. Which is most responsible?

10. How is the mass of the neutron determined?

11. The most stable nuclides have a mass number A near 60 (see Fig. 50-6). Why don't *all* nuclides have mass numbers near 60?

12. If we neglect the very lightest nuclides, the binding energy per nucleon in Fig. 50-6 is roughly constant at 7 to 8 MeV/nucleon. Do you expect the mean electronic binding energy per electron in atoms also to be roughly constant throughout the periodic table?

13. Why is the binding energy per nucleon (Fig. 50-6) low at low mass numbers? At high mass numbers?

14. In the binding energy curve of Fig. 50-6, what is special or notable about the nuclides ^{2}H, ^{4}He, ^{62}Ni, and ^{239}Pu?

15. The magnetic moment of the neutron is $-1.9130\ \mu_N$. What is a nuclear magneton and how does it differ from a Bohr magneton? What does the minus sign mean? How can the neutron, which carries no net charge, have a magnetic moment in the first place?

16. A particular ^{238}U nucleus was created in a massive stellar explosion, perhaps 10^{10} y ago. It suddenly decays by α emission while we are observing it. After all those years, why does it decide to decay at this particular moment?

17. Can you justify this statement: "In measuring half-lives by the method of Sample Problem 50-4, it is not necessary to measure the absolute decay rate R; any quantity proportional to it will suffice. However, in the method of Sample Problem 50-5 an absolute rate *is* needed."

18. Does the temperature affect the rate of decay of radioactive nuclides? If so, how?

19. You are running longevity tests on lightbulbs. Do you expect their "decay" to be exponential? What is the essential difference between the decay of lightbulbs and of radionuclides?

20. Generally clocks exhibit complete regularity of some periodic process. Considering that radioactive decay is completely random, how can it nevertheless be used for the measurement of time?

21. Can you give a justification, even a partial one, for the barrier-tunneling phenomenon in terms of basic ideas about the wave nature of matter?

22. Explain why, in alpha decay, short half-lives correspond to large disintegration energies, and conversely.

23. A radioactive nucleus can emit a positron, e$^+$. This corresponds to a proton in the nucleus being converted to a neutron. The mass of a neutron, however, is greater than that of a proton. How then can positron emission occur?

24. In beta decay the emitted electrons form a continuous spectrum, but in alpha decay the alpha particles form a discrete spectrum. What difficulties did this cause in the explanation of beta decay, and how were these difficulties finally overcome?

25. How do neutrinos differ from photons? Each has zero charge and (presumably) zero rest mass and travels at the speed of light.

26. The decay of radioactive elements produces helium, which eventually passes into the Earth's atmosphere. The amount of helium actually present in the atmosphere, however, is very much less than the amount released in this way. Explain.

27. The half-life of ^{238}U is 4.5×10^9 y, about the age of the solar system. How can such a long half-life be measured?

28. In radioactive dating with ^{238}U, how do you get around the fact that you do not know how much ^{238}U was present in the rocks to begin with? (*Hint:* What is the ultimate decay product of ^{238}U?)

29. Make a list of the various sources of ionizing radiation encountered in our environment, whether natural or artificial.

30. Which of these conservation laws apply to all nuclear reactions: conservation of (*a*) charge, (*b*) mass, (*c*) total energy, (*d*) rest energy, (*e*) kinetic energy, (*f*) linear momentum, (*g*) angular momentum, and (*h*) total number of nucleons?

31. Small temperature changes have a large effect on the rate of chemical reactions but generally have a negligible effect on the rate of nuclear reactions. Explain.

32. In the development of our understanding of the atom, did we use atomic models as we now use nuclear models? Is Bohr's theory such an atomic model? Are models now used in atomic physics? What is the difference between a model and a theory?

33. What are the basic assumptions of the collective and the independent particle models of nuclear structure? How do they differ? Are there similarities between them?

34. Does the collective model of the nucleus give us a picture of the following phenomena: (*a*) acceptance by the nucleus of a colliding particle, (*b*) loss of a particle by spontaneous emission, (*c*) fission, (*d*) dependence of stability on energy content?

35. What is so special ("magic") about the magic nucleon numbers?

36. Why are the magic nucleon numbers and the magic electron numbers not the same? What accounts for each?

37. The average number of stable (or very long-lived) isotopes of the inert gases is 3.7. The average number of stable nuclides for the four magic neutron numbers, however, is 5.8, considerably greater. If the inert gases are so stable, why were not more stable isotopes of them created when the elements were formed?

EXERCISES

50-1 Discovering the Nucleus

1. Calculate the distance of closest approach for a head-on collision between a 5.30-MeV α particle and the nucleus of a copper atom.

2. (*a*) Calculate the electric force on an α particle at the surface of a gold atom, presuming that the positive charge is spread uniformly throughout the volume of the atom. Ignore the atomic electrons. A gold atom has a radius of 0.16 nm; treat the α particle as a point particle. (*b*) Through what distance would this force, presumed constant, have to act to bring a 5.30-MeV α particle to rest? Express your answer in terms of the diameter of a gold atom.

3. Assume that a gold nucleus has a radius of 6.98 fm (see Table 50-1), and an α particle has a radius of 1.8 fm. What energy must an incident α particle have to just touch the gold nucleus?

50-2 Some Nuclear Properties

4. Locate the nuclides displayed in Table 50-1 on the nuclidic chart of Fig. 50-4. Which of these nuclides are within the stability zone?

5. The radius of a nucleus is measured, by electron-scattering methods, to be 3.6 fm. What is the likely mass number of the nucleus?

6. Arrange the 25 nuclides given here in squares as a section of the nuclidic chart similar to Fig. 50-4. Draw in and label (*a*) all isobaric (constant *A*) lines and (*b*) all lines of constant neutron excess, defined as $N - Z$. Consider nuclides $^{118-122}$Te, $^{117-121}$Sb, $^{116-120}$Sn, $^{115-119}$In, and $^{114-118}$Cd.

7. A neutron star is a stellar object whose density is about that of nuclear matter, as calculated in Sample Problem 50-2. Suppose that the Sun were to collapse into such a star without losing any of its present mass. What would be its expected radius?

8. Verify that the binding energy per nucleon given in Table 50-1 for ^{239}Pu is indeed 7.56 MeV/nucleon. The needed atomic masses are 239.052156 u (^{239}Pu), 1.007825 u (^{1}H), and 1.008665 u (neutron).

9. Calculate the average binding energy per nucleon of ^{62}Ni, which has an atomic mass of 61.928349 u. This nucleus has the greatest binding energy per nucleon of all the known stable nuclei.

10. The atomic masses of ^{1}H, ^{12}C, and ^{238}U are 1.007825 u, 12.000000 u (by definition), and 238.050783 u, respectively. (*a*) What would these masses be if the mass unit were defined so that the mass of ^{1}H was (exactly) 1.000000 u? (*b*) Use your result to suggest why this perhaps obvious choice was not made.

11. (*a*) Convince yourself that the energy tied up in nuclear, or strong-force, bonds is proportional to *A*, the mass number of the nucleus in question. (*b*) Convince yourself that the energy tied up in Coulomb-force bonds between the protons is proportional to $Z(Z - 1)$. (*c*) Show that, as we move to larger and larger nuclei (see Fig. 50-4), the importance of (*b*) increases more rapidly than does that of (*a*).

12. In the periodic table, the entry for magnesium is

12 Mg 24.305

There are three isotopes:

^{24}Mg, atomic mass = 23.985042 u;

^{25}Mg, atomic mass = 24.985837 u;

^{26}Mg, atomic mass = 25.982593 u.

The abundance of ^{24}Mg is 78.99% by mass. Calculate the abundances of the other two isotopes.

13. Because a nucleon is confined to a nucleus, we can take its uncertainty in position to be approximately the nuclear radius R. What does the uncertainty principle yield for the kinetic energy of a nucleon in a nucleus with, say, $A = 100$? (*Hint:* Take the uncertainty in momentum Δp to be the actual momentum p.)

14. You are asked to pick apart an α particle (^{4}He) by removing, in sequence, a proton, a neutron, and a proton. Calculate (*a*) the work required for each step, (*b*) the total binding energy of the α particle, and (*c*) the binding energy per nucleon. Needed atomic masses are

^{4}He	4.002603 u	^{2}H	2.014102 u
^{3}H	3.016049 u	^{1}H	1.007825 u
n	1.008665 u.		

15. To simplify calculations, atomic masses are sometimes tabulated, not as the actual atomic mass m but as $(m - A)c^2$, where A is the mass number expressed in mass units. This quantity, usually reported in MeV, is called the *mass excess,* symbol Δ. Using data from Sample Problem 50-3, find the mass excesses for (*a*) ^{1}H, (*b*) the neutron, and (*c*) ^{120}Sn.

16. (*a*) Show that the total binding energy of a nuclide can be written as

$$E_B = Z\Delta_H + N\Delta_n - \Delta,$$

where Δ_H, Δ_n, and Δ are the appropriate mass excesses; see Exercise 15. (*b*) Using this method calculate the binding energy per nucleon for ^{197}Au. Compare your result with the value listed in Table 50-1. The needed mass excesses are $\Delta_H = +7.289$ MeV, $\Delta_n = +8.071$ MeV, and $\Delta_{197} = -31.157$ MeV. Δ_H is the mass excess of ^{1}H. Note the economy of calculation that results when mass excesses are used in place of the actual masses.

17. A penny has a mass of 3.00 g. Calculate the nuclear energy that would be required to separate all the neutrons and protons in this coin. Ignore the binding energy of the electrons. For simplicity assume that the penny is made entirely of ^{63}Cu atoms (mass = 62.929601 u). The atomic masses of the proton and the neutron are 1.007825 u and 1.008665 u, respectively.

18. Nuclear radii may be measured by scattering high-energy electrons from nuclei. (*a*) What is the de Broglie wavelength for 480-MeV electrons? (*b*) Are they suitable probes for this purpose? Relativity must be taken into account.

50-3 Radioactive Decay

19. The half-life of a radioactive isotope is 140 d. How many days would it take for the activity of a sample of this isotope to fall to one-fourth of its initial decay rate?

20. The half-life of a particular radioactive isotope is 6.5 h. If there are initially 48×10^{19} atoms of this isotope in a particular sample, how many atoms of this isotope remain after 26 h?

21. A radioactive isotope of mercury, ^{197}Hg, decays into gold, ^{197}Au, with a decay constant of 0.0108 h^{-1}. (*a*) Calculate its half-life. (*b*) What fraction of the original amount will remain after three half-lives? (*c*) After 10 days?

22. From data presented in the first few paragraphs of Section 50-3, deduce (*a*) the disintegration constant λ and (*b*) the half-life of ^{238}U.

23. ^{67}Ga, atomic mass = 66.93 u, has a half-life of 78.25 h. Consider an initially pure 3.42-g sample of this isotope. (*a*) Find its activity (decay rate). (*b*) Find its activity 48.0 h later.

24. Show that the law of radioactive decay (Eq. 50-6) can be written in the form

$$N = N_0(\tfrac{1}{2})^{t/t_{1/2}}.$$

25. ^{223}Ra decays by alpha decay with a half-life of 11.43 d. How many helium atoms are created in 28 d from an initially pure sample of ^{223}Ra containing 4.70×10^{21} atoms?

26. The radionuclide ^{64}Cu has a half-life of 12.7 h. How much of an initially pure 5.50-g sample of ^{64}Cu will decay during the 2-h period beginning 14.0 h later?

27. The radionuclide ^{32}P (half-life = 14.28 d) is often used as a tracer to follow the course of biochemical reactions involving phosphorus. (*a*) If the counting rate in a particular experimental setup is 3050 counts/s, after what time will it fall to 170 counts/s? (*b*) A solution containing ^{32}P is fed to the root system of an experimental tomato plant and the ^{32}P activity in a leaf is measured 3.48 d later. By what factor must this reading be multiplied to correct for the decay that has occurred since the experiment began?

28. A 1.00-g sample of samarium emits α particles at a rate of 120 particles/s. ^{147}Sm, whose natural abundance in bulk samarium is 15.0%, is the responsible isotope. Calculate the half-life of this isotope.

29. ^{239}Pu, atomic mass = 239 u, decays by alpha decay with a half-life of 24,100 y. How many grams of helium are produced by an initially pure 12.0-g sample of ^{239}Pu after 20,000 y? (Recall that an α particle is a helium nucleus, with an atomic mass of 4.00 u.)

30. A source contains two phosphorus radionuclides, ^{32}P ($t_{1/2}$ = 14.3 d) and ^{33}P ($t_{1/2}$ = 25.3 d). Initially 10.0% of the decays come from ^{33}P. How long must one wait until 90.0% do so?

50-4 Alpha Decay

31. Consider a ^{238}U nucleus to be made up of an α particle (^{4}He) and a residual nucleus (^{234}Th). Plot the electrostatic potential energy $U(r)$, where r is the distance between these particles. Cover the range 10 fm $< r <$ 100 fm and compare your plot with that of Fig. 50-9.

32. Generally, heavier nuclides tend to be more unstable to alpha decay. For example, the most stable isotope of uranium, ^{238}U, has an alpha decay half-life of 4.5×10^9 y. The most stable isotope of plutonium is ^{244}Pu with a 8.2×10^7 y half-life, and for curium we have ^{248}Cm and 3.4×10^5 y. When half of an original sample of ^{238}U has decayed, what fractions of the original isotopes of (*a*) plutonium and (*b*) curium are left?

33. Heavy radionuclides emit an α particle rather than other combinations of nucleons because the α particle is such a stable, tightly bound structure. To confirm this, calculate the disintegration energies for these hypothetical decay processes and discuss the meaning of your findings:

$$^{235}\text{U} \rightarrow {}^{232}\text{Th} + {}^{3}\text{He}, \qquad Q_3;$$
$$^{235}\text{U} \rightarrow {}^{231}\text{Th} + {}^{4}\text{He}, \qquad Q_4;$$
$$^{235}\text{U} \rightarrow {}^{230}\text{Th} + {}^{5}\text{He}, \qquad Q_5.$$

The needed atomic masses are

^{232}Th	232.038050 u	^{3}He	3.016029 u
^{231}Th	231.036297 u	^{4}He	4.002603 u
^{230}Th	230.033127 u	^{5}He	5.012228 u.
^{235}U	235.043923 u		

34. Under certain circumstances, a nucleus can decay by emitting a particle heavier than an α particle. Such decays are very rare. Consider the decays

$$^{223}\text{Ra} \rightarrow {}^{209}\text{Pb} + {}^{14}\text{C}$$

and

$$^{223}\text{Ra} \rightarrow {}^{219}\text{Rn} + {}^{4}\text{He}.$$

(*a*) Calculate the Q-values for these decays and determine that both are energetically possible. (*b*) The Coulomb barrier height for α particles in this decay is 30 MeV. What is the barrier height for ^{14}C decay? Atomic masses are

^{223}Ra	223.018497 u	^{14}C	14.003242 u
^{209}Pb	208.981075 u	^{4}He	4.002603 u.
^{219}Rn	219.009475 u		

50-5 Beta Decay

35. ^{137}Cs is present in the fallout from above-ground detonations of nuclear bombs. Because it beta decays with a slow 30.2-y half-life into ^{137}Ba, releasing considerable energy in the process, it is an environmental concern. The atomic masses of the Cs and Ba are 136.907084 u and 136.905821 u, respectively. Calculate the total energy released in one decay.

36. A free neutron decays according to Eq. 50-16. Calculate the maximum energy $K_{\max}$ of the beta spectrum. Needed atomic masses are:

n	1.008665 u;	^{1}H	1.007825 u.

37. An electron is emitted from a middle-mass nuclide ($A = 150$, say) with a kinetic energy of 1.00 MeV. (*a*) Find its de Broglie wavelength. (*b*) Calculate the radius of the emitting nucleus. (*c*) Can such an electron be confined as a standing wave in a "box" of such dimensions? (*d*) Can you use these numbers to disprove the argument (long since abandoned) that electrons actually exist in nuclei?

38. The radionuclide ^{32}P decays to ^{32}S as described by Eq. 50-14. In a particular decay event, a 1.71-MeV electron is emitted, the maximum possible value. Find the kinetic energy of the recoiling ^{32}S atom in this event. The atomic mass of ^{32}S is 31.97 u. (*Hint:* For the electron it is necessary to use the relativistic expressions for the kinetic energy and the linear momentum. Newtonian mechanics may safely be used for the relatively slow-moving ^{32}S atom.)

50-6 Measuring Ionizing Radiation

39. The nuclide ^{198}Au, half-life = 2.693 d, is used in cancer therapy. Calculate the mass of this isotope required to produce an activity of 250 Ci.

40. A Geiger counter records 8722 counts in 1 min. Calculate the activity of the source in Ci, assuming that the counter records all decays.

41. A typical chest x-ray radiation dose is 25 mrem, delivered by x rays with a quality factor of 0.85. Assuming that the mass of the exposed tissue is one-half the patient's mass of 88 kg, calculate the energy absorbed in joules.

42. A 75-kg person receives a whole-body radiation dose of 24 mrad, delivered by α particles for which the quality factor is 12. Calculate (*a*) the absorbed energy in joules and (*b*) the equivalent dose in rem.

43. An activity of 3.94 μCi is needed in a radioactive sample to be used in a medical procedure. One week before treatment, a nuclide sample with a half-life of 1.82×10^5 s is prepared. What should be the activity of the sample at the time of preparation in order that it have the required activity at the time of treatment?

44. The plutonium isotope ^{239}Pu, atomic mass 239.05 u, is produced as a by-product in nuclear reactors and hence is accumulating in reactor fuel elements. It is radioactive, decaying by alpha decay with a half-life of 2.411×10^4 y. However, plutonium is also one of the most toxic chemicals known; as little as 2.00 mg is lethal to a human. (*a*) How many nuclei constitute a chemically lethal dose? (*b*) What is the decay rate of this amount? (*c*) Its activity in curies?

45. Cancer cells are more vulnerable to x and gamma radiation than are healthy cells. Though linear accelerators are now replacing it, in the past the standard source for radiation therapy has been radioactive ^{60}Co, which beta decays into an excited nuclear state of ^{60}Ni, which immediately drops into the ground state, emitting two gamma-ray photons, each of approximate energy 1.2 MeV. The controlling beta-decay half-life is 5.27 y. How many radioactive ^{60}Co nuclei are present in a 6000-Ci source used in a hospital? The atomic mass of ^{60}Co is 59.93 u.

46. An airline pilot spends an average of 20 h per week flying at 12,000 m, at which altitude the dose equivalent rate due to cosmic and solar radiation is 12 μSv/h (1 Sv = 1 sievert = 100 rem; the sievert is the SI unit of dose equivalent). Calculate the annual equivalent dose in mrem.

47. After long effort, in 1902, Marie and Pierre Curie succeeded in separating from uranium ore the first substantial quantity of radium, 1 decigram (dg) of pure RaCl_2. The radium was the radioactive isotope ^{226}Ra, which decays with a half-life of 1600 y. (*a*) How many radium nuclei had they isolated? (*b*) What was the decay rate of their sample, in Bq? (1 Bq = 1 becquerel = 1 decay/s.) (*c*) In curies? The molar mass of Cl is 35.453 g/mol; the atomic mass of the radium isotope is 226.03 u.

48. Calculate the mass of 4.60 μCi of ^{40}K, which has a half-life of 1.28×10^9 y and an atomic mass of 40.0 u.

50-7 Natural Radioactivity

49. A rock is found to contain 4.20 mg of ^{238}U and 2.00 mg of ^{206}Pb. Assume that the rock contained no lead at formation, with all the lead now present arising from the decay of the uranium. Find the age of the rock. The half-life of ^{238}U is 4.47×10^9 y.

50. A particular rock is thought to be 260 million years old. If it contains 3.71 mg of ^{238}U, how much ^{206}Pb should it contain?

51. A rock, recovered from far underground, is found to contain 860 μg of ^{238}U, 150 μg of ^{206}Pb, and 1.60 mg of ^{40}Ca. How much ^{40}K will it very likely contain? Needed half-lives are 4.47×10^9 y for ^{238}U and 1.28×10^9 y for ^{40}K.

50-8 Nuclear Reactions

52. Fill in the missing nuclide in each of the following reactions: (*a*) ^{116}Sn(?,p)^{117}Sn; (*b*) ^{40}Ca(α,n)?; and (*c*) ?(p,n)^{7}Be.

53. Calculate Q for the reaction ^{59}Co(p,n)^{59}Ni. Needed atomic masses are

^{59}Co	58.933200 u	^{1}H	1.007825 u
^{59}Ni	58.934352 u	n	1.008665 u.

54. Making mental use of the overlay of Fig. 50-13 applied to Fig. 50-12, write down the reactions by which the radionuclide ^{197}Pt ($t_{1/2} = 18.3$ h) can be prepared, at least in principle. Except in special circumstances, only stable nuclides can serve as practical targets for nuclear reactions.

55. The radionuclide ^{60}Co ($t_{1/2} = 5.27$ y) is much used in cancer therapy. Tabulate possible reactions that might be used in preparing it. Limit the projectiles to neutrons, protons, and deuterons. Limit the targets to stable nuclides. The stable nuclides suitably close to ^{60}Co are ^{63}Cu, ^{60,61,62}Ni, ^{59}Co, and ^{57,58}Fe. (Commercially, ^{60}Co is made by bombarding elemental cobalt, which consists only of the isotope ^{59}Co, with neutrons in a reactor.)

56. A beam of deuterons falls on a copper target. Copper has two stable isotopes, ^{63}Cu (69.2%) and ^{65}Cu (30.8%). Tabulate the residual nuclides that can be produced by the reactions (d,n), (d,p), (d,α), and (d,γ). By inspection of Fig. 50-4, indicate which residual nuclides are stable and which are radioactive.

57. Prepare an overlay like that of Fig. 50-13 in which that figure is extended to include reactions involving the light nuclides ^{3}H (tritium) and ^{3}He, considered both as projectiles and as emerging particles.

58. A platinum target is bombarded with cyclotron-accelerated deuterons for several hours and then the element iridium ($Z = 77$) is separated chemically from it. What radioisotopes of iridium are present and by what reactions are they formed? (*Note:* ^{190}Pt and ^{192}Pt, not shown in Fig. 50-12, are stable platinum isotopes, but their isotopic abundances are so small that we may ignore their presence.)

59. Prepare an overlay like that of Fig. 50-13 in which two nucleons or light nuclei may appear as emerging particles. The reaction ^{63}Cu(α,pn)^{65}Zn is an example. Consider the combinations nn, np, and pd as possibilities.

50-9 Nuclear Models

60. From the following list of nuclides, identify (*a*) those with filled nucleon shells, (*b*) those with one nucleon outside a filled shell, and (*c*) those with one vacancy in an otherwise filled shell. The nuclides are ^{13}C, ^{18}O, ^{40}K, ^{49}Ti, ^{60}Ni, ^{91}Zr, ^{92}Mo, ^{121}Sb, ^{143}Nd, ^{144}Sm, ^{205}Tl, and ^{207}Pb.

61. The nucleus ^{91}Zr ($Z = 40$, $N = 51$) has a single neutron outside a filled 50-neutron core. Because 50 is a magic number, this neutron should perhaps be especially loosely bound. (*a*) Calculate its binding energy. (*b*) Calculate the binding energy of the next neutron, which must be extracted from the filled core. (*c*) Find the binding energy per particle for the entire nucleus. Compare these three numbers and discuss. Needed atomic masses are

^{91}Zr	90.905645 u	n	1.008665 u
^{90}Zr	89.904704 u	^{1}H	1.007825 u.
^{89}Zr	88.908889 u		

PROBLEMS

1. When an α particle collides elastically with a nucleus, the nucleus recoils. A 5.00-MeV α particle has a head-on elastic collision with a gold nucleus, initially at rest. What is the kinetic energy (*a*) of the recoiling nucleus and (*b*) of the rebounding α particle? The mass of the α particle may be taken to be 4.00 u and that of the gold nucleus to be 197 u.

2. Because the neutron has no charge, its mass must be found in some way other than by using a mass spectrometer. When a resting neutron and a proton meet, they combine and form a deuteron, emitting a gamma ray whose energy is 2.2233 MeV. The atomic masses of the proton and the deuteron are 1.007825 u and 2.014102 u, respectively. Find the mass of the neutron from these data, to as many significant figures as the data warrant.

3. The spin and the magnetic moment (maximum z component) of ^{7}Li in its ground state (see Table 50-1) are $\frac{3}{2}$ and $+3.26$ nuclear magnetons, respectively. A free ^{7}Li nucleus is placed in a magnetic field of 2.16 T. (*a*) Into how many levels will the ground state split because of space quantization? (*b*) What is the energy difference between adjacent pairs of levels? (*c*) What is the wavelength that corresponds to a transition between such a pair of levels? (*d*) In what region of the electromagnetic spectrum does this wavelength lie?

4. (*a*) Show that the electrostatic potential energy of a uniform sphere of charge Q and radius R is given by

$$U = \frac{3Q^2}{20\pi\epsilon_0 R}.$$

(*Hint:* Assemble the sphere from thin spherical shells brought in from infinity.) (*b*) Find the electrostatic potential energy for the nuclide ^{239}Pu, assumed spherical; see Table 50-1. (*c*) Compare its electrostatic potential energy per particle with its binding energy per nucleon of 7.56 MeV. (*d*) What do you conclude?

5. A certain radionuclide is being manufactured in a cyclotron, at a constant rate P. It is also decaying, with a disintegration constant λ. Let the production process continue for a time that is long compared to the half-life of the radionuclide. Show that the number of radioactive nuclei present at such times will be constant and will be given by $N = P/\lambda$. Show further that this result holds no matter how many of the radioactive nuclei were present initially. The nuclide is said to be in *secular equilibrium* with its source; in this state its decay rate is equal to its production rate.

6. The radionuclide ^{56}Mn has a half-life of 2.58 h and is produced in a cyclotron by bombarding a manganese target with deuterons. The target contains only the stable manganese isotope ^{55}Mn, and the reaction that produces ^{56}Mn is

$$^{55}\text{Mn} + \text{d} \rightarrow {}^{56}\text{Mn} + \text{p}.$$

After being bombarded for a time $\gg 2.58$ h, the activity of the target, due to ^{56}Mn, is $8.88 \times 10^{10}\ \text{s}^{-1}$; see Problem 5. (*a*) At what constant rate P are ^{56}Mn nuclei being produced in the cyclotron during the bombardment? (*b*) At what rate are they decaying (also during the bombardment)? (*c*) How many ^{56}Mn nuclei are present at the end of the bombardment? (*d*) What is their total mass? The atomic mass of ^{56}Mn is 55.94 u.

7. A radium source contains 1.00 mg of ^{226}Ra, which decays with a half-life of 1600 y to produce ^{222}Rn, an inert gas. This radon gas in turn decays by alpha decay with a half-life of 3.82 d. (*a*) Calculate the decay rate of ^{226}Ra in the source. (*b*) At what rate is the radon decaying when it has come to secular equilibrium with the radium source? See Problem 5. (*c*) How much radon is in secular equilibrium with the radium source?

8. There is speculation that the free proton may not actually be a stable particle but may be radioactive, with a half-life of about 1×10^{32} y. If this turns out to be true, about how long would you have to wait to be reasonably sure that one proton in your body has decayed? Assume that you are made of water and have a mass of 70 kg.

9. A ^{238}U nucleus emits an α particle of energy 4.196 MeV. Calculate the disintegration energy Q for this process, taking the recoil energy of the residual ^{234}Th nucleus into account. The atomic mass of an α particle is 4.0026 u and that of the ^{234}Th is 234.04 u. Compare your result with that of Sample Problem 50-6*a*.

10. Consider that a ^{238}U nucleus emits (*a*) an α particle or (*b*) a sequence of neutron, proton, neutron, proton. Calculate the energy released in each case. (*c*) Convince yourself both by reasoned argument and also by direct calculation that the difference between these two numbers is just the total binding energy of the α particle. Find that binding energy. Needed atomic masses are

^{238}U	238.050783 u	^{4}He	4.002603 u
^{237}U	237.048724 u	^{1}H	1.007825 u
^{236}Pa	236.048674 u	n	1.008665 u.
^{235}Pa	235.045432 u		
^{234}Th	234.043596 u		

11. The radionuclide ^{11}C decays according to

$$^{11}\text{C} \rightarrow {}^{11}\text{B} + e^{+} + \nu, \qquad t_{1/2} = 20.3 \text{ min.}$$

The maximum energy of the positron spectrum is 960.8 keV. (*a*) Show that the disintegration energy Q for this process is given by

$$Q = (m_{\text{C}} - m_{\text{B}} - 2m_e)c^2,$$

where m_{C} and m_{B} are the atomic mass of ^{11}C and ^{11}B, respectively and m_e is the electron (positron) mass. (*b*) Given that $m_{\text{C}} = 11.011434$ u, $m_{\text{B}} = 11.009305$ u, and $m_e = 0.0005486$ u, calculate Q and compare it with the maximum energy of the positron spectrum, given above. (*Hint:* Let m'_{C} and m'_{B} be the nuclear masses and proceed as in Sample Problem 50-7 for beta decay. Note that positron decay is an exception to the general rule that, if atomic masses are used in nuclear decay processes, the mass of the emitted electron is automatically taken care of.)

12. Some radionuclides decay by capturing one of their own atomic electrons, a *K*-electron, say. An example is

$$^{49}\text{V} + e^{-} \rightarrow {}^{49}\text{Ti} + \nu, \qquad t_{1/2} = 331 \text{ d.}$$

(*a*) Show that the disintegration energy Q for this process is given by

$$Q = (m_{\text{V}} - m_{\text{Ti}})c^2 - E_{\text{K}},$$

where m_{V} and m_{Ti} are the atomic masses of ^{49}V and ^{49}Ti, respectively, and E_{K} is the binding energy of the vanadium *K*-electron. (*Hint:* Put m'_{V} and m'_{Ti} as the corresponding nuclear masses and proceed as in Sample Problem 50-7; see the footnote in that sample problem.) (*b*) Find the disintegration energy Q for the decay of ^{49}V by *K*-electron capture. The needed data are $m_{\text{V}} = 48.948517$ u, $m_{\text{Ti}} = 48.947871$ u, and $E_K = 5.47$ keV.

13. One of the dangers of radioactive fallout from a nuclear bomb is ^{90}Sr, which beta decays with a 29-y half-life. Because it has chemical properties much like calcium, the strontium, if eaten by a cow, becomes concentrated in its milk and ends up in the bones of whoever drinks the milk. The energetic decay electrons damage the bone marrow and thus impair the production of red blood cells. A 1-megaton bomb produces approximately 400 g of ^{90}Sr. If the fallout spreads uniformly over a 2000-km^2 area, what area would have radioactivity equal to the allowed bone burden for one person of 0.002 mCi? The atomic mass of ^{90}Sr is 89.9 u.

14. An 87-kg worker at a breeder reactor plant accidentally ingests 2.5 mg of ^{239}Pu dust. ^{239}Pu has a half-life of 24,100 y, decaying by alpha decay. The energy of the emitted α particles is 5.2 MeV, with a quality factor of 13. Assume that the plutonium resides in the worker's body for 12 h, and that 95% of the emitted α particles are stopped within the body. Calculate (*a*) the number of plutonium atoms ingested, (*b*) the number that decay during the 12 h, (*c*) the energy absorbed by the body, (*d*) the resulting physical dose in rad, and (*e*) the equivalent biological dose in rem.

15. Two radioactive materials that are unstable to alpha decay, ^{238}U and ^{232}Th, and one that is unstable to beta decay. ^{40}K, are sufficiently abundant in granite to contribute significantly to the heating of the Earth through the decay energy produced. The alpha-unstable isotopes give rise to decay chains that stop at stable lead isotopes. ^{40}K has a single beta decay. Decay information follows:

Parent Nuclide	*Decay Mode*	*Half-life* (y)	*Stable Endpoint*	*Q* (MeV)	*f* (ppm)
^{238}U	α	4.47×10^{9}	^{206}Pb	51.7	4
^{232}Th	α	1.41×10^{10}	^{208}Pb	42.7	13
^{40}K	β	1.28×10^{9}	^{40}Ca	1.32	4

Q is the total energy released in the decay of one parent nucleus to the final stable endpoint and f is the abundance of the isotope in kilograms per kilogram of granite; ppm means parts per million. (*a*) Show that these materials give rise to a total heat production of 987 pW for each kilogram of granite. (*b*) Assuming that there is 2.7×10^{22} kg of granite in a 20-km thick, spherical shell around the Earth, estimate the power this will produce over the whole Earth. Compare this with the total solar power intercepted by the Earth, 1.7×10^{17} W.

16. Consider the reaction $X(a,b)Y$, in which X is taken to be at rest in the laboratory reference frame. The initial kinetic energy in this frame is

$$K_{\text{lab}} = \tfrac{1}{2} m_a v_a^2.$$

(*a*) Show that the initial velocity of the center of mass of the system in the laboratory frame is

$$V = v_a \left(\frac{m_a}{m_X + m_a} \right).$$

Is this quantity changed by the reaction? (*b*) Show that the initial kinetic energy, viewed now in a reference frame attached to the center of mass of the two particles, is given by

$$K_{\text{cm}} = K_{\text{lab}}\left(\frac{m_X}{m_X + m_a}\right).$$

Is this quantity changed by the reaction? (*c*) In the reaction ^{90}Zr(d,p)^{91}Zr the kinetic energy of the deuteron, measured in the laboratory frame, is 15.9 MeV. Find $v_a\ (= v_{\text{d}})$, V, and K_{cm}. Ignore the small relativistic effects.

17. In an endothermic reaction ($Q < 0$), the interacting particles a and X must have a kinetic energy, measured in the center-of-mass reference frame, of at least $|Q|$ if the reaction is to "go." Show, using the result of Problem 16, that the *threshold energy* for particle a, measured in the laboratory reference frame, is

$$K_{\text{th}} = |Q|\frac{m_X + m_a}{m_X}.$$

Is it reasonable that K_{th} should be greater than $|Q|$?

18. The nuclide ^{208}Pb is "doubly magic" in that both its proton number $Z\ (=82)$ and its neutron number $N\ (=126)$ represent filled nucleon shells. An additional proton would yield ^{209}Bi and an additional neutron ^{209}Pb. These "extra" nucleons should be easier to remove than a proton or a neutron from the filled shells of ^{208}Pb. (*a*) Calculate the energy required to remove the "extra" proton from ^{209}Bi and compare it with the energy required to remove a proton from the filled proton shell of ^{208}Pb. (*b*) Calculate the energy required to remove the "extra" neutron from ^{209}Pb and compare it with the energy required to remove a neutron from the filled neutron shell of ^{208}Pb. Do your results agree with expectation? Use these atomic mass data:

Nuclide	*Z*	*N*	*Atomic Mass* (u)
^{209}Bi	82 + 1	126	208.980383
^{208}Pb	82	126	207.976636
^{207}Tl	82 − 1	126	206.977408
^{209}Pb	82	126 + 1	208.981075
^{207}Pb	82	126 − 1	206.975881

The atomic masses of the proton and the neutron are 1.007825 u and 1.008665 u, respectively.

COMPUTER PROBLEM

1. After a brief neutron irradiation of silver, two activities are present: ^{108}Ag ($t_{1/2} = 2.42$ min) with an initial decay rate of 3.1×10^5/s, and ^{110}Ag ($t_{1/2} = 24.6$ s) with an initial decay rate of 4.1×10^6/s. Make a plot similar to Fig. 50-8 showing the total combined decay rate of the two isotopes as a function of time from $t = 0$ until $t = 10$ min. In Fig. 50-8, the extraction of the half-life for simple decays was illustrated. Given only the plot of total decay rate, can you suggest a way to analyze it in order to find the half-lives of both isotopes?

CHAPTER 51

ENERGY FROM THE NUCLEUS

In a system of interacting particles, we can extract useful energy when the system moves to a lower energy state (that is, a more tightly bound state). In an atomic system, we can extract this energy through chemical reactions, such as burning. In a nuclear system, we can extract energy in a variety of ways. For example, the energy released in radioactive decay has been used to provide electrical power to cardiac pacemakers and to space probes.

In this chapter we consider the two primary methods that are used to extract energy from the nucleus and convert it to useful purposes. In nuclear fission, *a heavy nucleus is split into two fragments. In* nuclear fusion, *two light nuclei are combined into a heavier nucleus. Figure 50-6 showed that either of these processes can result in more tightly bound nuclei and therefore can release excess nuclear binding energy to be converted into other forms of energy. Reactors based on nuclear fission today provide a significant share of the world's electrical power. Research and engineering are actively under way to develop reactors based on nuclear fusion.*

51-1 THE ATOM AND THE NUCLEUS

When we get energy from coal by burning it in a furnace, we are tinkering with *atoms* of carbon and oxygen, rearranging their outer *electrons* in more stable combinations. When we get energy from uranium by consuming it in a nuclear reactor, we are tinkering with its *nucleus,* rearranging its *nucleons* in more stable combinations.

Electrons are held in atoms by the Coulomb force, and it takes a few electron volts to remove one of the outer electrons. On the other hand, nucleons are held in nuclei by the strong nuclear force, and it takes a few *million* electron volts to pull one of *them* out. This factor is also reflected in our ability to extract about a million times more energy from a kilogram of uranium than from a kilogram of coal.

In both the atomic and nuclear cases, the appearance of energy is accompanied by a decrease in the rest energy of the fuel. The only difference between consuming uranium and burning coal is that, in the former case, a much larger fraction of the available rest energy (again, by a factor of several million) is converted to other forms of energy.

We must be clear about whether our concern is for the quantity of energy or for the rate at which the energy is delivered—that is, the *power.* In the nuclear case, will the kilogram of uranium burn slowly in a power reactor or explosively in a bomb? In the atomic case, are we thinking about exploding a stick of dynamite or digesting a jelly doughnut? (Surprisingly, the energy release is greater in the second case than in the first!)

Table 51-1 shows how much energy can be extracted from 1 kg of matter by doing various things to it. Instead of reporting the energy directly, we measure it by showing how long the extracted energy could operate a 100-W lightbulb. Row 5, the total mutual annihilation of matter and antimatter, is the ultimate in extracting energy from matter. When you have used up all the available mass you can do no more. (However, no one has yet figured out an economi-

TABLE 51-1 **Energy from 1 kg of Matter**

Form of Matter	*Process*	*Time*[a]
Water	A 50-m waterfall	5 s
Coal	Burning	8 h
^{235}U	Fission	3×10^4 y
Hot deuterium gas	Fusion	3×10^4 y
Matter and antimatter	Annihilation	3×10^7 y

[a] These numbers show how long the energy generated could power a 100-W lightbulb.

cal way to produce and store 1 kg of antimatter to use for energy production.)

Keep in mind that the comparisons of Table 51-1 are on a per-unit-mass basis. Kilogram for kilogram we get several million times more energy from uranium than we do from coal or from falling water. On the other hand, there is a lot of coal in the Earth's crust and a lot of water backed up behind the Bonneville Dam on the Columbia River.

51-2 NUCLEAR FISSION: THE BASIC PROCESS

In 1932, the English physicist James Chadwick discovered the neutron. A few years later Enrico Fermi and his collaborators in Rome discovered that, if various elements are bombarded by these new projectiles, new radioactive elements are produced. Fermi had predicted that the neutron, being uncharged, would be a useful nuclear projectile; unlike the proton or the α particle, it experiences no repulsive Coulomb force when it approaches a nuclear surface. Because there is no Coulomb barrier for it, the slowest neutron can penetrate and interact with even the most massive, highly charged nucleus. *Thermal neutrons,* which are neutrons in equilibrium with matter at room temperature, are convenient and effective bombarding particles. At 300 K, the average kinetic energy of such neutrons is

$$K_{\text{av}} = \tfrac{3}{2}kT = \tfrac{3}{2}(8.62 \times 10^{-5}\ \text{eV/K})(300\ \text{K}) = 0.04\ \text{eV}.$$

In 1939, the German chemists Otto Hahn and Fritz Strassmann, following work initiated by Fermi and his collaborators, bombarded uranium with thermal neutrons. They found by chemical analysis that after the bombardment a number of new radioactive elements were present, among them one whose chemical properties were remarkably similar to those of barium. Repeated tests finally convinced these chemists that this "new" element was not new at all; it really *was* barium. How could this middle-mass element ($Z = 56$) be produced by bombarding uranium ($Z = 92$) with neutrons?

The riddle was solved within a few weeks by physicists Lise Meitner and her nephew Otto Frisch. They showed that a uranium nucleus, having absorbed a neutron, could split, with the release of energy, into two roughly equal parts, one of which might well be barium. They named this process *nuclear fission.**

The fission of ^{235}U by thermal neutrons, a process of great practical importance, can be represented by

$$^{235}\text{U} + \text{n} \rightarrow X + Y + b\text{n}. \qquad (51\text{-}1)$$

Here X and Y stand for *fission fragments,* middle-mass nuclei that are usually highly radioactive. The factor b, which has the average value 2.47 for fission events of this type, is the number of neutrons released in such events.

Figure 51-1 shows two visible tracks made by fission fragments X and Y in a process such as is represented by Eq. 51-1. Two details are immediately apparent: (1) The two tracks do not have the same length. This occurs because the two fragments do not have the same mass; when the uranium nucleus splits, the two pieces tend to have different masses. Figure 51-2 shows the mass distribution; as you can see, there is a high probability that one fragment will have a mass number of about 95 and the other about 140. You can also see that the chance of an even split is only about 1 in 10^4. (2) The second apparent feature of Fig. 51-1 is that the two tracks are not quite exactly back to back. Can you explain this?

A heavy nucleus such as ^{235}U has a neutron-to-proton ratio of about 1.6. On the average, we expect that the fragments X and Y should have about the same neutron-to-proton ratio. A glance at Fig. 50-4 shows that, for example, a stable nucleus with mass number 95 has about 40 protons and 55 neutrons, for a neutron-to-proton ratio of about 1.4. If the $A = 95$ fragment formed in fission has a neutron-

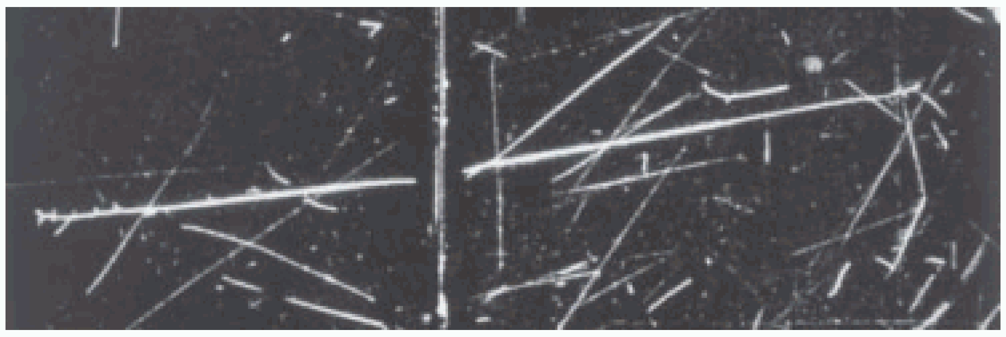


FIGURE 51-1. When a fast charged particle passes through a cloud chamber, it leaves a track of liquid droplets. The two back-to-back tracks represent fission fragments, produced by a fission event that took place in a thin vertical uranium foil in the center of the chamber.

---

*See "The Discovery of Fission," by Otto Frisch and John Wheeler, *Physics Today,* November 1967, p. 43. Also see *Lise Meitner, A Life in Physics* by Ruth Lewin Sime (University of California Press, 1996).

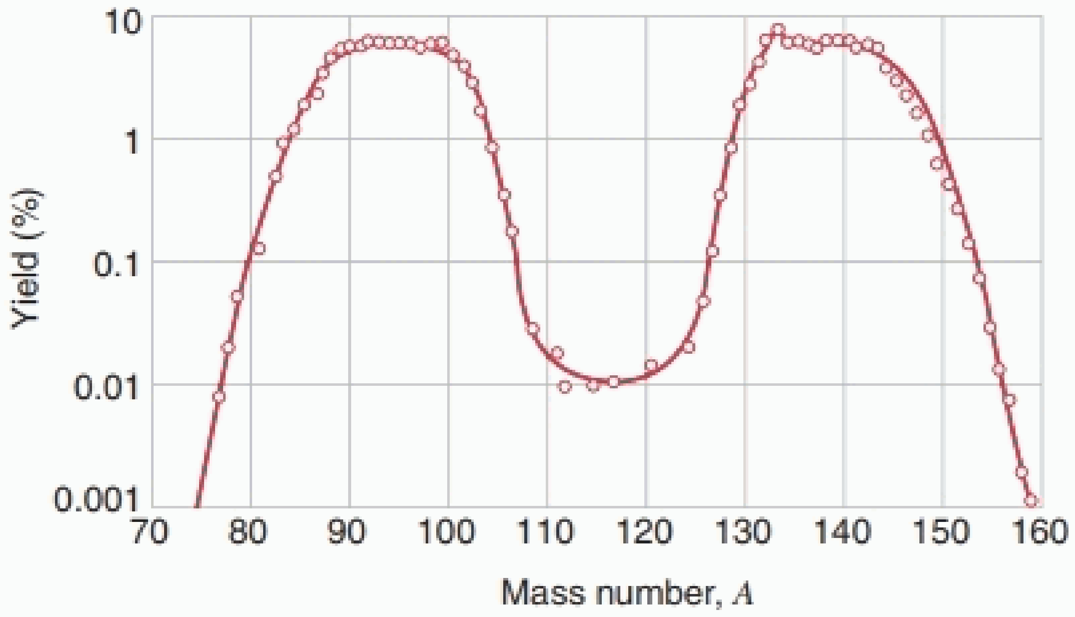


FIGURE 51-2. The distribution in mass of the fission fragments X and Y (see Eq. 51-1) from the fission of ^{235}U by thermal neutrons. Note that the vertical scale is logarithmic.

to-proton ratio of 1.6, it will be far from the line of stable nuclei. This explains why the fission fragments are radioactive. Usually through successive beta decays (and sometimes by neutron emission), they try to relieve their neutron excess and move toward the region of stable nuclei.

Let us consider a specific example of a fission process. When ^{235}U captures a neutron, it forms for an instant an unstable, highly excited nucleus of ^{236}U that can then fission, perhaps as follows:

$$^{235}\text{U} + \text{n} \rightarrow {}^{236}\text{U} \rightarrow {}^{140}\text{Xe} + {}^{94}\text{Sr} + 2\text{n}. \qquad (51\text{-}2)$$

The fission fragments ^{140}Xe and ^{94}Sr decay until each reaches a stable end product, as follows:

$$^{140}\text{Xe} \xrightarrow[14\text{ s}]{} {}^{140}\text{Cs} \xrightarrow[65\text{ s}]{} {}^{140}\text{Ba} \xrightarrow[13\text{ d}]{} {}^{140}\text{La} \xrightarrow[40\text{ h}]{} {}^{140}\text{Ce (stable)}$$

$$^{94}\text{Sr} \xrightarrow[75\text{ s}]{} {}^{94}\text{Y} \xrightarrow[19\text{ min}]{} {}^{94}\text{Zr (stable)}$$

The decays are β^- events, the half-lives being indicated at each stage. As for all beta decays, the mass numbers (140 and 94) remain unchanged as the decays continue.

The disintegration energy Q for fission is very much larger than for chemical processes. We can support this by a rough calculation. From the binding energy curve of Fig. 50-6, we see that for heavy nuclides ($A = 240$, say) the binding energy per nucleon is about 7.6 MeV. In the intermediate range ($A = 120$, say), it is about 8.5 MeV. The difference in total binding energy between a single nucleus ($A = 240$) and two fragments (assumed equal) into which it may be split is then

$$Q = 2(8.5\text{ MeV})\frac{A}{2} - (7.6\text{ MeV})A \approx 200\text{ MeV}.$$

Sample Problem 51-1 shows a more careful calculation, which agrees very well with this rough estimate.

SAMPLE PROBLEM 51-1. Calculate the disintegration energy Q for the fission event of Eq. 51-2, taking into account the decay of the fission fragments. Needed atomic masses are

^{235}U	235.043923 u	^{140}Ce	139.905434 u
n	1.008665 u	^{94}Zr	93.906316 u.

Solution If we replace the fission fragments in Eq. 51-2 by their stable end products, we see that the overall transformation is

$$^{235}\text{U} \rightarrow {}^{140}\text{Ce} + {}^{94}\text{Zr} + \text{n}.$$

The single neutron comes about because the (initiating) neutron on the left side of Eq. 51-2 cancels one of two neutrons on the right side of that equation.

The mass difference $\Delta m = m_\text{i} - m_\text{f}$ for this reaction is

$$\begin{aligned}\Delta m &= 235.043923\text{ u} - (139.905434\text{ u} + 93.906316\text{ u} \\ &\quad + 1.008665\text{ u}) \\ &= 0.223508\text{ u},\end{aligned}$$

and the corresponding energy is

$$Q = \Delta m\, c^2 = (0.223508\text{ u})(931.5\text{ MeV/u}) = 208.2\text{ MeV},$$

in good agreement with our previous rough estimate of 200 MeV.

About 80% of the disintegration energy is in the form of the kinetic energy of the two fragments, the remainder going to the neutron and the radioactive decay products.

If the fission event takes place in a bulk solid, most of the disintegration energy appears as an increase in the internal energy of the solid, which shows a corresponding rise in temperature. Five percent or so of the disintegration energy, however, is associated with neutrinos that are emitted during the beta decay of the primary fission fragments. This energy is carried out of the system and does not contribute to the increase in its internal energy.

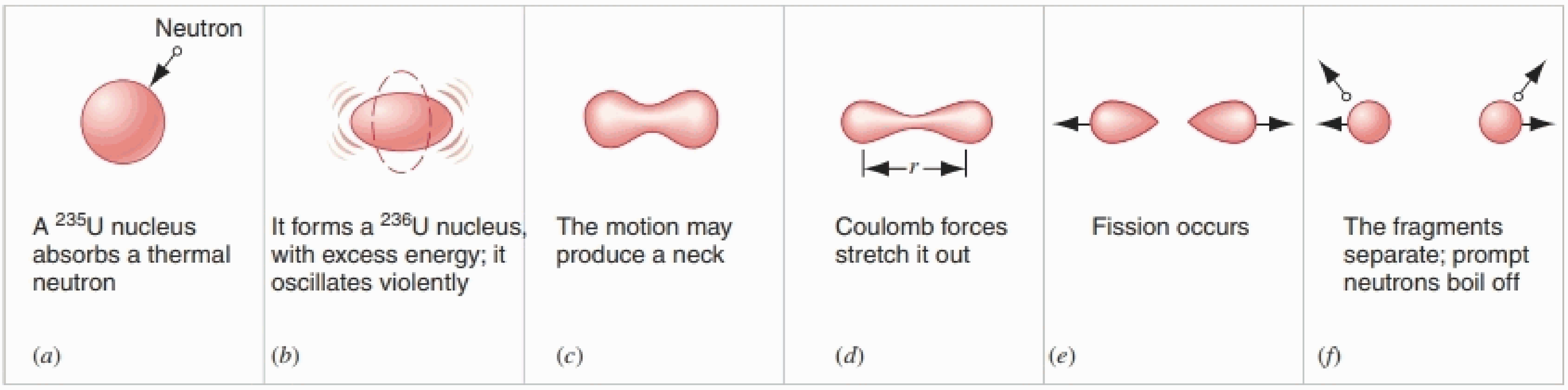
51-3 THEORY OF NUCLEAR FISSION

Soon after the discovery of fission, Niels Bohr and John Wheeler developed a theory, based on the analogy between a nucleus and a charged liquid drop, that explained its main features.

Figure 51-3 suggests how the fission process proceeds. When a heavy nucleus such as ^{235}U absorbs a slow neutron, as in Fig. 51-3*a*, that neutron falls into the potential well associated with the strong nuclear forces that act in the nuclear interior. Its potential energy is then transformed into internal excitation energy, as Fig. 51-3*b* suggests. As we show in Sample Problem 51-2, the resulting excitation energy of ^{236}U is considerable, about 6.5 MeV.

Figure 51-3*c* shows that the nucleus, behaving like an energetically oscillating, charged liquid drop, will sooner or later develop a short "neck" and will begin to separate into two charged "globs." If conditions are right, the electrostatic repulsion between these two globs will force them apart, breaking the neck. The two fragments, each still carrying some residual excitation energy, then fly apart. Fission has occurred.

So far this model gives a good qualitative picture of the fission process. It remains to be seen, however, whether it can answer a hard question: Why are some heavy nuclides (^{235}U and ^{239}Pu, say) readily fissionable by slow neutrons but other, equally heavy, nuclides (^{238}U and ^{243}Am, say) are not?



A ^{235}U nucleus absorbs a thermal neutron (*a*)

It forms a ^{236}U nucleus, with excess energy; it oscillates violently (*b*)

The motion may produce a neck (*c*)

Coulomb forces stretch it out (*d*)

Fission occurs (*e*)

The fragments separate; prompt neutrons boil off (*f*)

FIGURE 51-3. The stages in a fission process, according to the liquid-drop fission model.

Bohr and Wheeler were able to answer this question. Figure 51-4 shows the potential energy curve for the fission process that they derived from their model. The horizontal axis displays the *distortion parameter r,* which is a rough measure of the extent to which the oscillating nucleus departs from a spherical shape. Figure 51-3*d* suggests how this parameter is defined before fission occurs. When the fragments are far apart, this parameter is simply the distance between their centers.

The energy interval between the initial state and the final state of the fissioning nucleus—that is, the disintegration energy Q—is displayed in Fig. 51-4. The central feature of that figure, however, is that the potential energy curve passes through a maximum at a certain value of r. There is a *potential barrier* of height E_b that must be surmounted (or tunneled) before fission can occur. This reminds us of alpha decay (see Fig. 50-9), which also is a process that is inhibited by a potential barrier. We see then that fission will occur *only* if the absorbed neutron provides an excitation energy E_n great enough to overcome the barrier or to have a reasonable probability of tunneling through it.

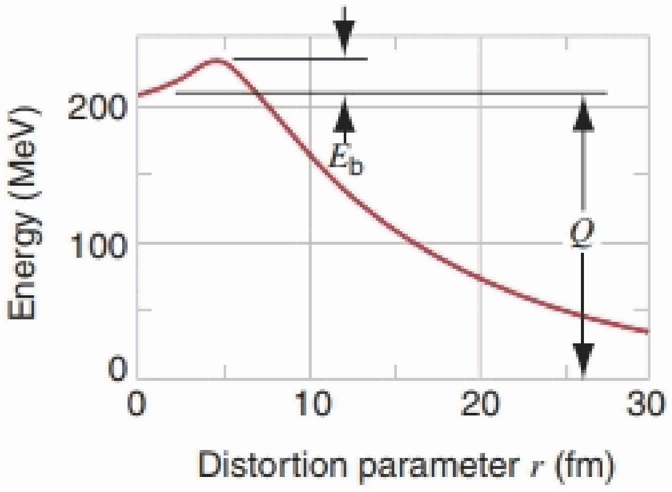


FIGURE 51-4. The potential energy at various stages in the fission process, showing the disintegration energy Q and the barrier height E_b.

Table 51-2 shows a test of fissionability by thermal neutrons applied to four heavy nuclides, chosen from dozens of candidates that might have been considered. For each nuclide both the barrier height E_b and the excitation energy E_n are given. E_b was calculated from the theory of Bohr and Wheeler, and E_n was computed (as in Sample Problem 51-2) from the known masses.

For ^{235}U and ^{239}Pu we see that $E_n > E_b$. This means that fission by absorbing a thermal neutron is predicted to occur for these nuclides. This is confirmed by noting, in the table, the large measured cross sections (that is, the reaction probabilities) for the process.

For the other two nuclides (^{238}U and ^{243}Am), we have $E_n < E_b$, so that there is not enough energy to surmount the barrier or to tunnel through it effectively. The excited nucleus (Fig. 51-3*b*) prefers to get rid of its excitation energy by emitting a gamma ray instead of by breaking into two large fragments. Table 51-2 shows, as we expect, that the cross sections for thermal neutron fission in these cases are exceedingly small. These nuclides *can* be made to fission, however, if they absorb a substantially energetic (rather than a thermal) neutron. For ^{238}U, for example, the absorbed neutron must have an energy of at least 1.3 MeV for the fission process to "go" with reasonable probability.

TABLE 51-2 **Test of the Fissionability of Four Nuclides**

Target Nuclide	*Nuclide Being Fissioned*	E_b *(MeV)*	E_n *(MeV)*	$E_n - E_b$ *(MeV)*	*Fission Cross Section*[a] *(barns)*
^{235}U	^{236}U	5.2	6.5	+1.3	584
^{238}U	^{239}U	5.7	4.8	−0.6	5×10^{-6}
^{239}Pu	^{240}Pu	4.8	6.4	+1.6	742
^{243}Am	^{244}Am	5.8	5.5	−0.3	<0.08

[a] The cross section is a measure of the probability for a nuclear reaction to occur. The cross section is measured in units of barns, where 1 barn = 10^{-28} m^2.

SAMPLE PROBLEM 51-2. Consider a ^{236}U nucleus in its ground state. How much energy is required to remove a neutron from it, leaving a ^{235}U nucleus behind? The needed atomic masses are

^{235}U 235.043923 u; n 1.008665 u; ^{236}U 236.045562 u.

Solution The increase in mass of the system as the neutron is pulled out is

$$\begin{aligned}\Delta m &= 1.008665 \text{ u} + 235.043923 \text{ u} - 236.045562 \text{ u}\\ &= 0.007026 \text{ u}.\end{aligned}$$

This means that an energy equal to

$$E_n = \Delta m\, c^2 = (0.007026 \text{ u})(931.5 \text{ MeV/u}) = 6.545 \text{ MeV}$$

must be expended. This, by definition, is the binding energy of the neutron in the ^{236}U nucleus.

When a ^{235}U nucleus absorbs a thermal neutron, 6.545 MeV is the amount of excitation energy that the thermal neutron brings into the ^{236}U nucleus. In effect, the ^{236}U nucleus is formed in an excited state 6.545 MeV above the ground state. The excited nucleus can get rid of this energy either by emitting gamma rays (which leaves a ^{236}U nucleus in its ground state) or by fission (see Eq. 51-1). It turns out that fission is about six times more likely than gamma-ray emission.

51-4 NUCLEAR REACTORS: THE BASIC PRINCIPLES

Energy releases per atom in individual nuclear events such as alpha emission are roughly a million times larger than those of chemical events. To make large-scale use of nuclear energy, we must arrange for one nuclear event to trigger another until the process spreads throughout bulk matter like a flame through a burning log. The fact that more neutrons are generated in fission than are consumed (see Eq. 51-1) raises just this possibility; the neutrons that are produced can cause fission in nearby nuclei and in this way a chain of fission events can propagate itself. Such a process is called a *chain reaction.* It can either be rapid and uncontrolled as in a nuclear bomb or controlled as in a nuclear reactor.

Suppose that we wish to design a nuclear reactor based, as most present reactors are, on the fission of ^{235}U by slow neutrons. The fuel in such reactors is almost always artificially "enriched," so that ^{235}U makes up a few percent of the uranium nuclei rather than the 0.7% that occurs in natural uranium; the remaining 99.3% of natural uranium is ^{238}U, which is not fissionable by thermal neutrons. Although on the average 2.47 neutrons are produced in ^{235}U fission for every thermal neutron consumed, there are serious difficulties in making a chain reaction "go." Here are three of the difficulties, together with their solutions:

1. *The neutron leakage problem.* A certain percentage of the neutrons produced will simply leak out of the reactor core and be lost to the chain reaction. If too many do so, the reactor will not work. Leakage is a surface effect, its magnitude proportional to the *square* of a typical reactor core dimension (surface area = $4\pi r^2$ for a sphere). Neutron production, however, is a volume effect, proportional to the *cube* of a typical dimension (volume = $\frac{4}{3}\pi r^3$ for a sphere). The fraction of neutrons lost by leakage can be made as small as we wish by making the reactor core large enough, thereby decreasing its surface-to-volume ratio (= $3/r$ for a sphere).

2. *The neutron energy problem.* Fission produces *fast* neutrons, with kinetic energies of about 2 MeV, but fission is induced most effectively by *slow* neutrons. The fast neutrons can be slowed down by mixing the uranium fuel with a substance that has these properties: (*a*) it is effective in causing neutrons to lose kinetic energy by collisions and (*b*) it does not absorb neutrons excessively, thereby removing them from the fission chain. Such a substance is called a *moderator.* Most power reactors in this country are moderated by water, in which the hydrogen nuclei (protons) are the effective moderating element.

3. *The neutron capture problem.* Neutrons may be captured by nuclei in ways that do not result in fission. The most common possibility is capture followed by the emission of a gamma ray. In particular, as the fast (MeV) neutrons generated in the fission processes are slowed down in the moderator to thermal equilibrium (0.04 eV), they must pass through an energy interval (1 − 100 eV) in which they are particularly susceptible to nonfission capture by ^{238}U.

To minimize such *resonance capture,* as it is called, the uranium fuel and the moderator (water, say) are not intimately mixed but are "clumped," remaining in close contact with each other but occupying different regions of the reactor volume. The hope is that a fast fission neutron, produced in a uranium "clump" (which might be a fuel rod), will with high probability find itself in the moderator as it passes through the "dangerous" resonance energy range. Once it has reached thermal energies, it will very likely wander back into a clump of fuel and produce a fission event. The task for reactor designers is to produce the most effective geometrical arrangement of fuel and moderator.

Figure 51-5 shows the neutron balance in a typical power reactor operating with a steady output. Let us trace the behavior of a sample of 1000 thermal neutrons in the reactor core. They produce 1330 neutrons by fission in the ^{235}U fuel and 40 more by fast fission in the ^{238}U, making a total of 370 new neutrons, all of them fast. Exactly this same number of neutrons is then lost to the chain by leakage from the core and by nonfission capture, leaving 1000 thermal neutrons to continue the chain. What has been gained in this cycle, of course, is that each of the 370 neutrons produced by fission has deposited about 200 MeV of energy in the reactor core, heating it up.

An important reactor parameter is the *multiplication factor k,* the ratio of the number of neutrons present at the beginning of a particular generation to the number present at the beginning of the next generation. For the situation of

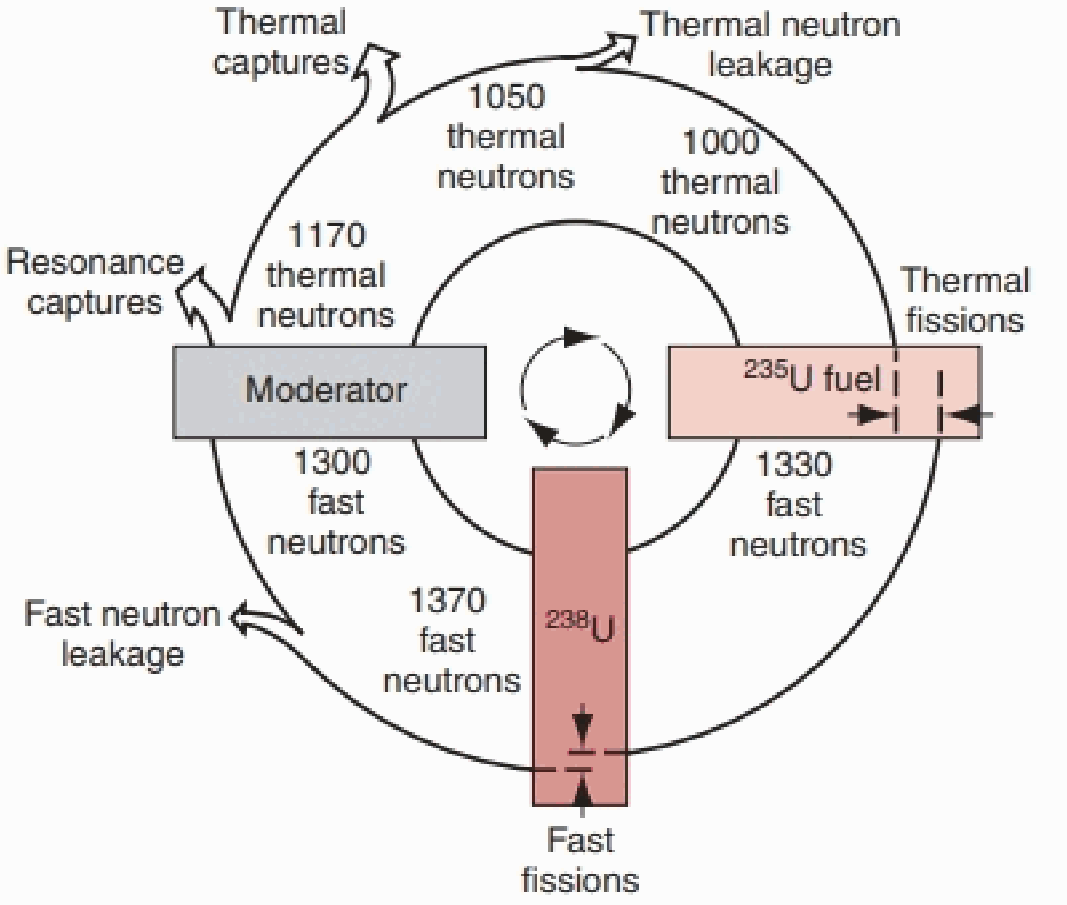

Thermal captures
Thermal neutron leakage
1050 thermal neutrons
1000 thermal neutrons
1170 thermal neutrons
Resonance captures
Thermal fissions
Moderator
235U fuel
1300 fast neutrons
1330 fast neutrons
1370 fast neutrons
238U
Fast neutron leakage
Fast fissions

FIGURE 51-5. A generation of 1000 thermal neutrons is followed through various stages in a reactor. At a steady operating level, the loss of neutrons due to captures (in the fuel, moderator, and structural elements) and leakage through the surface is exactly balanced by the production of neutrons in the fission processes.

Fig. 51-5, the multiplication factor is exactly 1. For $k = 1$, the operation of the reactor is said to be exactly *critical,* which is what we wish it to be for steady power production. Reactors are designed so that they are inherently *supercritical* ($k > 1$); the multiplication factor is then adjusted to critical operation ($k = 1$) by inserting *control rods* into the reactor core. These rods, containing a material such as cadmium that absorbs neutrons readily, can then be withdrawn as needed to compensate for the tendency of reactors to go subcritical as (neutron-absorbing) fission products build up in the core during continued operation.

If you pulled out one of the control rods, how fast would the reactor power level increase? This *response time* is controlled by the fascinating circumstance that a small fraction of the neutrons generated by fission is not emitted promptly from the newly formed fission fragments but is emitted from these fragments later, as they decay by beta emission. Of the 370 "new" neutrons analyzed in Fig. 51-5, for example, about 16 are delayed, being emitted from fragments following beta decays whose half-lives range from 0.2 to 55 s. These delayed neutrons are few in number but they serve the useful purpose of slowing down the reactor response time to match human reaction times.

Figure 51-6 shows the broad outlines of an electric power plant based on a *pressurized-water reactor* (PWR), a type in common use in the United States. In such a reactor, water is used both as the moderator and as the heat transfer medium. In the *primary loop,* water at high temperature and pressure (possibly 600 K and 150 atm) circulates through the reactor vessel and transfers heat from the reactor core to the steam generator, which provides high-pressure steam to operate the turbine that drives the generator. To complete the *secondary loop,* low-pressure steam from the turbine is condensed to water and forced back into the steam generator by a pump. To give some idea of scale, a typical reactor vessel for a 1000-MW (electric) plant may be 10 m high and weigh 450 tons. Water flows through the primary loop at a rate of about 300,000 gal/min.

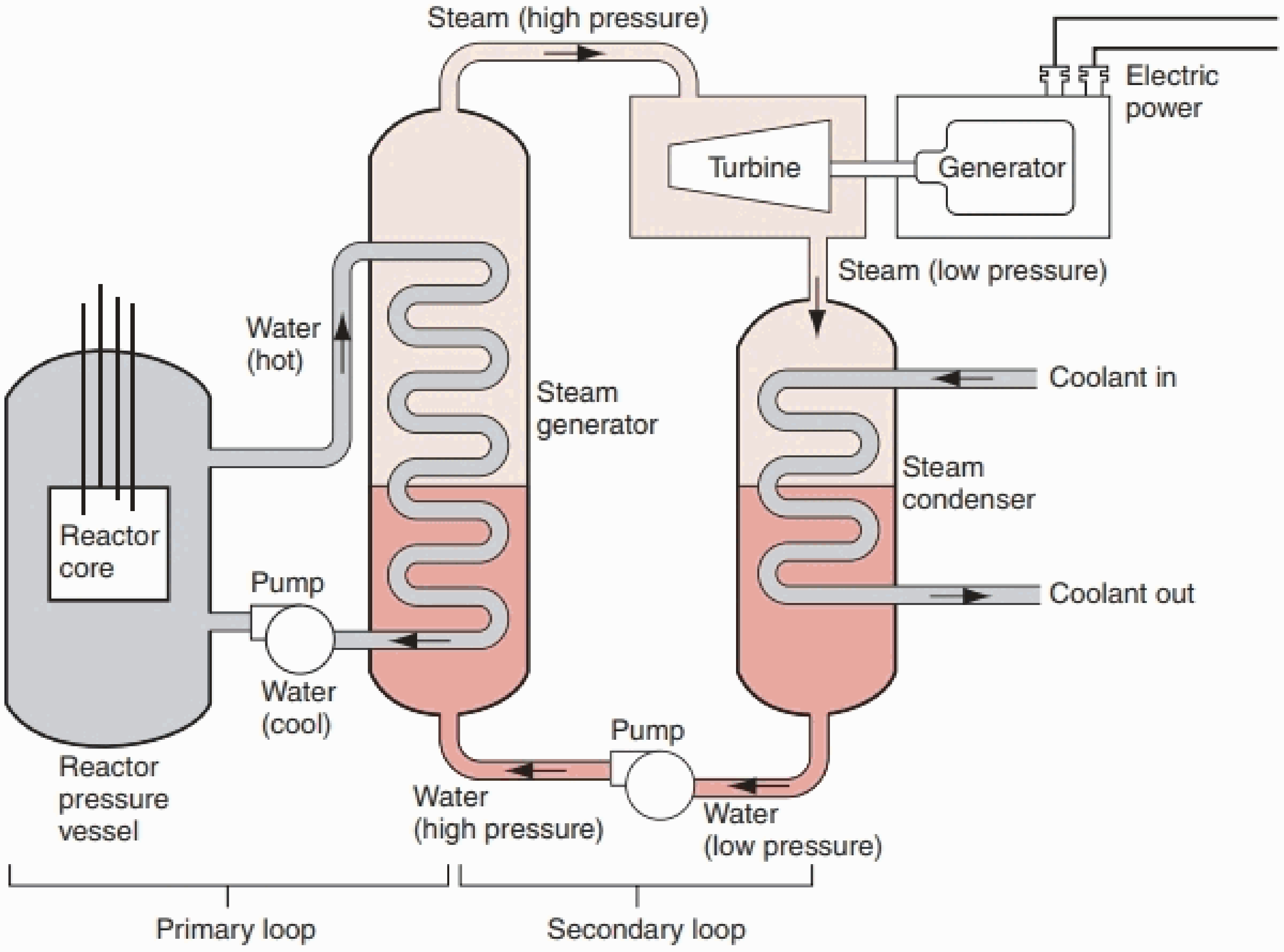

Steam (high pressure)
Electric power
Turbine
Generator
Steam (low pressure)
Water (hot)
Steam generator
Coolant in
Steam condenser
Reactor core
Coolant out
Pump
Water (cool)
Pump
Reactor pressure vessel
Water (high pressure)
Water (low pressure)
Primary loop
Secondary loop

FIGURE 51-6. A simplified layout of a nuclear power plant based on a pressurized-water reactor.

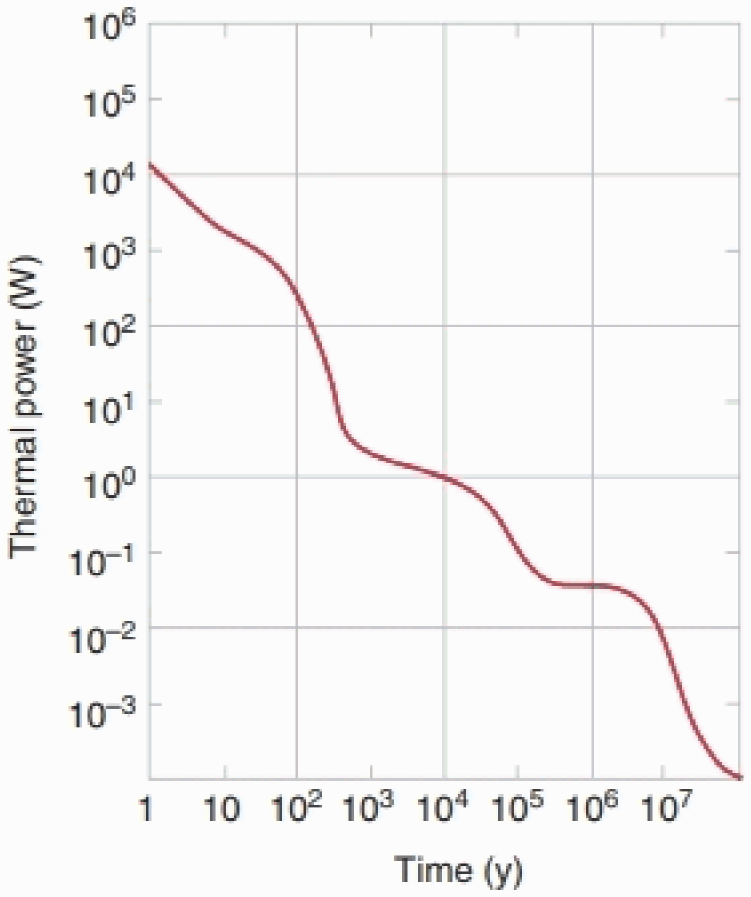


FIGURE 51-7. Thermal power released by the radioactive wastes of one year's operation of a typical large nuclear power plant, as a function of time after the fuel is removed. The curve represents the effect of many radionuclides with a range of half-lives. Note that both scales are logarithmic.

An unavoidable feature of reactor operation is the accumulation of radioactive wastes, including both fission products and heavy "transuranic" nuclides such as plutonium and americium. One measure of their radioactivity is the rate at which they release energy in thermal form. Figure 51-7 shows the variation with time of the thermal power generated by such wastes from one year's operation of a typical large nuclear plant. Note that both scales are logarithmic. The total activity of the waste 10 years after its removal from the reactor is about 3×10^7 Ci.

SAMPLE PROBLEM 51-3. A large electric generating station is powered by a pressurized-water nuclear reactor. The thermal power in the reactor core is 3400 MW, and 1100 MW of electricity is generated. The fuel consists of 86,000 kg of uranium, in the form of 110 tons of uranium oxide, distributed among 57,000 fuel rods. The uranium is enriched to 3.0% ^{235}U. (*a*) What is the plant efficiency? (*b*) At what rate *R* do fission events occur in the reactor core? (*c*) At what rate is the ^{235}U fuel disappearing? Assume conditions at start-up? (*d*) At this rate of fuel consumption, how long would the fuel supply last? (*e*) At what rate is mass being lost in the reactor core?

Solution (*a*) The efficiency *e* is the ratio between the power output (in the form of electric energy) to the power input (in the form of thermal energy), or

$$e = \frac{\text{electric output}}{\text{thermal input}} = \frac{1100\ \text{MW}}{3400\ \text{MW}} = 0.32 \text{ or } 32\%.$$

As for all power plants, whether based on fossil fuel or nuclear fuel, the efficiency is controlled by the second law of thermodynamics. In this plant, 3400 MW − 1100 MW or 2300 MW of power must be discharged as thermal energy to the environment. (*b*) If *P* (= 3400 MW) is the thermal power in the core and *Q* (= 200 MeV) is the average energy released per fission event, then, in steady-state operation,

$$R = \frac{P}{Q} = \frac{3.4 \times 10^9\ \text{J/s}}{(200\ \text{MeV/fission})(1.6 \times 10^{-13}\ \text{J/MeV})}$$
$$= 1.06 \times 10^{20}\ \text{fissions/s}.$$

(*c*) ^{235}U disappears by fission at the rate calculated in (*b*). It is also consumed by (nonfission) neutron capture at a rate about one-fourth as large. The total ^{235}U consumption rate is then $(1.25)(1.06 \times 10^{20}\ \text{s}^{-1})$ or $1.33 \times 10^{20}\ \text{s}^{-1}$. We recast this as a mass rate as follows:

$$\frac{dM}{dt} = (1.33 \times 10^{20}\ \text{s}^{-1})\left(\frac{0.235\ \text{kg/mol}}{6.02 \times 10^{23}\ \text{atoms/mol}}\right)$$
$$= 5.19 \times 10^{-5}\ \text{kg/s} = 4.5\ \text{kg/d}.$$

(*d*) From the data given, we can calculate that, at start-up, about (0.03)(86,000 kg) or 2600 kg of ^{235}U were present. Thus a somewhat simplistic answer would be

$$T = \frac{2600\ \text{kg}}{4.5\ \text{kg/d}} = 580\ \text{d}.$$

In practice, the fuel rods are replaced (often in batches) before their ^{235}U content is entirely consumed.
(*e*) From Einstein's $E = \Delta m\, c^2$ relation, we can write

$$\frac{dM}{dt} = \frac{dE/dt}{c^2} = \frac{3.4 \times 10^9\ \text{W}}{(3.00 \times 10^8\ \text{m/s})^2}$$
$$= 3.8 \times 10^{-8}\ \text{kg/s} = 3.3\ \text{g/d}.$$

The mass loss rate is about the mass of one penny every day! This mass loss rate (reduction in rest energy) is quite a different quantity than the fuel consumption rate (loss of ^{235}U) calculated in part (*c*).

51-5 A NATURAL REACTOR

On December 2, 1942, when the reactor assembled by Enrico Fermi and his associates first went critical, they had every right to expect that they had put into operation the first fission reactor that had ever existed on this planet. About 30 years later it was discovered that, if they did in fact think that, they were wrong.

Some two billion years ago, in a uranium deposit now being mined in Gabon, West Africa, a natural fission reactor went into operation and ran for perhaps several hundred thousand years before shutting itself off.

The story of this discovery is fascinating at the level of the best detective thriller. More important, it provides a first-class example of the nature of the scientific evidence needed to back up what may seem at first to be an improbable claim. It set a high standard for all who speculate about past events. We consider here only two points.*

---

*For the complete story, see "A Natural Fission Reactor," by George A. Cowan, *Scientific American*, July 1976, p. 36.

1. *Was there enough fuel?* The fuel for a uranium-based fission reactor must be the easily fissionable isotope ^{235}U, which constitutes only 0.72% of natural uranium. This isotopic ratio has been measured not only for terrestrial samples but also in Moon rocks and in meteorites, in which the same value is always found. The initial clue to the discovery in Gabon was that the uranium from this deposit was deficient in ^{235}U, some samples having an abundance as low as 0.44%. Investigation led to the speculation that this deficit in ^{235}U could be accounted for if, at some time in the past, this isotope was partially consumed by the operation of a natural fission reactor.

The serious problem remains that, with an isotopic abundance of only 0.72%, a reactor can be assembled (as Fermi and his team learned) only with the greatest of difficulty. There seems no chance at all that it could have happened naturally.

However, things were different in the distant past. Both ^{235}U and ^{238}U are radioactive, with half-lives of 0.704×10^9 y and 4.47×10^9 y, respectively. Thus the half-life of the readily fissionable ^{235}U is about 6.4 times shorter than that of ^{238}U. Because ^{235}U decays faster, there must have been more of it, relative to ^{238}U, in the past. Two billion years ago, in fact, this abundance was not 0.72%, as it is now, but 3.8%. This abundance happens to be just about the abundance to which natural uranium is artificially enriched to serve as fuel in modern power reactors.

With this amount of readily fissionable fuel available in the distant past, the presence of a natural reactor (providing certain other conditions are met) is much less surprising. The fuel was there. Two billion years ago, incidentally, the highest order of life forms that had evolved were the blue-green algae.

2. *What is the evidence?* The mere depletion of ^{235}U in an ore deposit is not enough evidence on which to base a claim for the existence of a natural fission reactor. More convincing proof is needed.

If there were a reactor, there must also be fission products; see Fig. 51-2. Of the 30 or so elements whose stable isotopes are produced in this way, some must still remain. Study of their isotopic ratios could provide the convincing evidence we need.

Of the several elements investigated, the case of neodymium is spectacularly convincing. Figure 51-8*a* shows the isotopic abundances of the seven stable neodymium isotopes as they are normally found in nature. Figure 51-8*b* shows these abundances as they appear among the ultimate stable products of the fission of ^{235}U. The clear differences are not surprising, considering their totally different origins. The isotopes shown in Fig. 51-8*a* were formed in supernova explosions that occurred before the formation of our solar system. The isotopes of Fig. 51-8*b* were cooked up in a reactor by totally different processes. Note particularly that ^{142}Nd, the dominant isotope in the natural element, is totally absent from the fission products.

The big question is: What do the neodymium isotopes found in the uranium ore body in Gabon look like? We must expect that, if a natural reactor operated there, isotopes from *both* sources (that is, natural isotopes as well as fission-produced isotopes) might be present. Figure 51-8*c* shows the results after this and other corrections have been made to the raw data. Comparison of Figs. 51-8*b* and 51-8*c* certainly suggests that there was indeed a natural fission reactor at work!

SAMPLE PROBLEM 51-4. The isotopic ratio of ^{235}U to ^{238}U in natural uranium deposits today is 0.0072. What was this ratio 2.0×10^9 y ago? The half-lives of the two isotopes are 0.704×10^9 y and 4.47×10^9 y, respectively.

Solution Consider two samples that, at a time t in the past, contained $N_5(0)$ and $N_8(0)$ atoms of ^{235}U and ^{238}U, respectively. The numbers of atoms remaining at the present time are

$$N_5(t) = N_5(0)e^{-\lambda_5 t} \quad \text{and} \quad N_8(t) = N_8(0)e^{-\lambda_8 t},$$

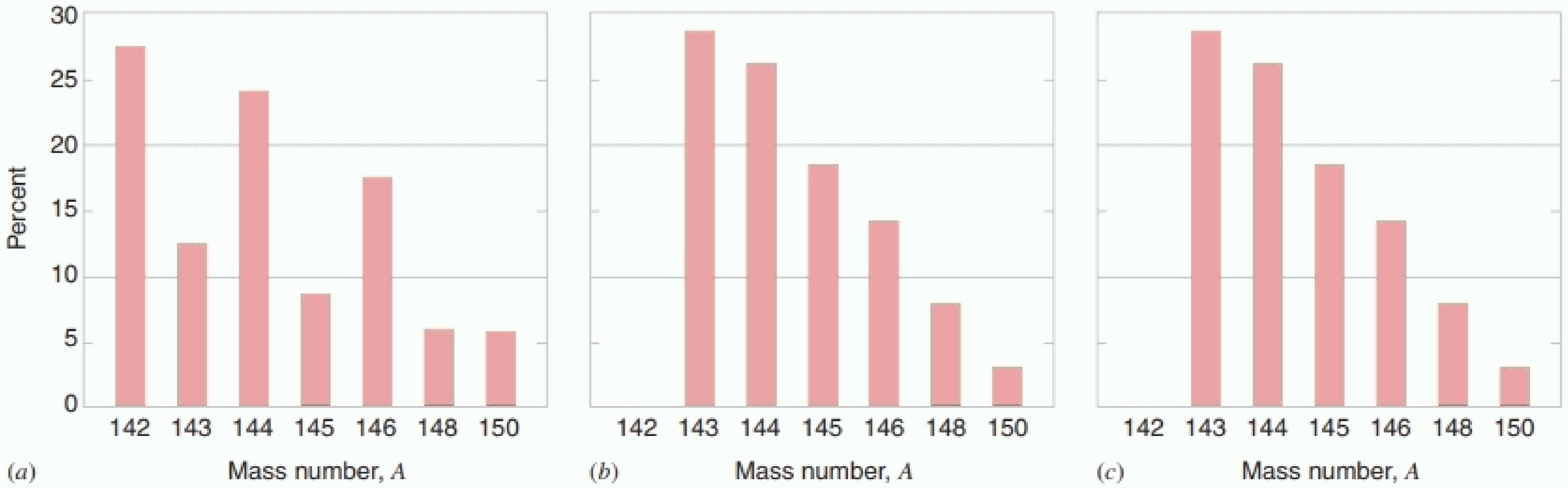


FIGURE 51-8. The distribution by mass number of the isotopes of neodymium as they occur in (*a*) natural terrestrial deposits, (*b*) the spent fuel of a power reactor, and (*c*) the uranium mine in Gabon, West Africa. Note that (*b*) and (*c*) are virtually identical and quite different from (*a*).

respectively, in which λ_5 and λ_8 are the corresponding disintegration constants. Dividing gives

$$\frac{N_5(t)}{N_8(t)} = \frac{N_5(0)}{N_8(0)}\, e^{-(\lambda_5 - \lambda_8)t}.$$

Expressed in terms of the isotopic ratio $R = N_5/N_8$, this becomes

$$R(0) = R(t)e^{(\lambda_5 - \lambda_8)t}.$$

The disintegration constants are related to the half-lives by Eq. 50-8, or

$$\lambda_5 = \frac{\ln 2}{t_{1/2}(^{235}\mathrm{U})} = \frac{0.693}{7.04 \times 10^8\ \mathrm{y}} = 0.984 \times 10^{-9}\ \mathrm{y}^{-1}$$

and

$$\lambda_8 = \frac{\ln 2}{t_{1/2}(^{238}\mathrm{U})} = \frac{0.693}{4.47 \times 10^9\ \mathrm{y}} = 0.155 \times 10^{-9}\ \mathrm{y}^{-1}.$$

Substituting in the expression for the isotopic ratio gives

$$\begin{aligned} R(0) &= R(t)e^{(\lambda_5 - \lambda_8)t} \\ &= (0.0072)e^{(0.984-0.155)(10^{-9}\ \mathrm{y}^{-1})(2.00\times 10^9\ \mathrm{y})} \\ &= (0.0072)e^{1.65} = 0.0378 \text{ or } 3.78\%. \end{aligned}$$

We see that, two billion years ago, the ratio of ^{235}U to ^{238}U in natural uranium deposits was much higher than it is today. When the Earth was formed (4.5 billion years ago) this ratio was 30%.

51-6 THERMONUCLEAR FUSION: THE BASIC PROCESS

We pointed out in connection with the binding energy curve of Fig. 50-6 that energy can be released if light nuclei are combined to form nuclei of somewhat larger mass number, a process called *nuclear fusion.* However, this process is hindered by the mutual Coulomb repulsion that tends to prevent two such (positively) charged particles from coming within range of each other's attractive nuclear forces and "fusing." This reminds us of the potential barrier that inhibits nuclear fission (see Fig. 51-4) and also of the barrier that inhibits alpha decay (see Fig. 50-9).

In the case of alpha decay, two charged particles—the α particle and the residual nucleus—are initially *inside* their mutual potential barrier. For alpha decay to occur, the α particle must leak through this barrier by the barrier-tunneling process and appear on the *outside.* In nuclear fusion the situation is reversed. Here the two particles must penetrate their mutual barrier from the *outside* if a nuclear interaction is to occur.

The interaction between two deuterons is of particular importance in fusion. Sample Problem 51-5 gives a rough calculation of the potential barrier between two deuterons, which works out to be about 200 keV. The corresponding barrier for two interacting ^{3}He nuclei (charge $= +2e$) is about 1 MeV. For more highly charged particles the barrier, of course, is correspondingly higher.

One way to arrange for light nuclei to penetrate their mutual Coulomb barrier is to use one light particle as a target and to accelerate the other by means of a cyclotron or a similar device. To generate power in a useful way from the fusion process, however, we must have the interaction of matter in bulk, just as in the combustion of coal. The cyclotron technique holds no promise in this direction. The best hope for obtaining fusion in bulk matter in a controlled fashion is to raise the temperature of the material so that the particles have sufficient energy to penetrate the barrier due to their thermal motions alone. This process is called *thermonuclear fusion.*

The average thermal kinetic energy K_{av} of a particle in equilibrium at a temperature T is given, as we have seen in Section 22-4, by

$$K_{\mathrm{av}} = \tfrac{3}{2}kT, \tag{51-3}$$

where k ($= 8.62 \times 10^{-5}$ eV/K) is the Boltzmann constant. At room temperature ($T \approx 300$ K), $K_{\mathrm{av}} = 0.04$ eV, which is, of course, far too small for our purpose.

Even at the center of the Sun, where $T \approx 1.5 \times 10^7$ K, the mean thermal kinetic energy calculated from Eq. 51-3 is only 1.9 keV. This still seems hopelessly small in view of the magnitude of the Coulomb barrier of 200 keV calculated in Sample Problem 51-5. Yet we know that thermonuclear fusion not only occurs in the solar interior but is its central and dominant feature.

The puzzle is solved with the realization that (1) the energy calculated from Eq. 51-3 is a *mean* kinetic energy; particles with energies much greater than this mean value constitute the high-energy "tails" of the Maxwellian speed distribution curves (see Fig. 22-6). Also, (2) the barrier heights that we have quoted represent only the *peaks* of the barriers. Barrier tunneling can occur to a significant extent at energies well below these peaks, as we saw in Section 50-4 in the case of alpha decay.

Figure 51-9 summarizes the situation by a quantitative example. The curve marked $n(K)$ in this figure is a Maxwell distribution of kinetic energies (see Section 22-5) drawn to correspond to the Sun's central temperature,

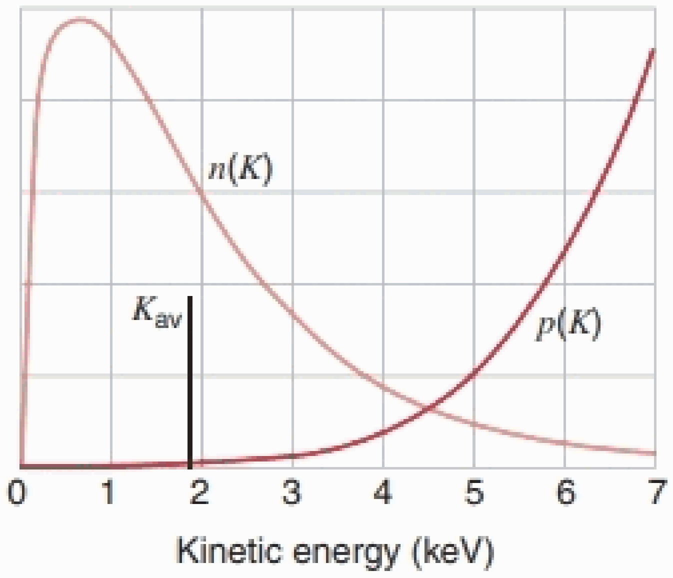


FIGURE 51-9. The curve marked $n(K)$ gives the distribution in energy of protons in the core of the Sun, corresponding to a temperature of 1.5×10^7 K. The vertical line indicates the mean kinetic energy per particle at that temperature. The curve marked $p(K)$ gives the probability of barrier penetration in proton–proton collisions. The two curves are drawn to different arbitrary vertical scales.

1.5×10^7 K. Although the same curve holds no matter what particle is under consideration, we focus our attention on protons, bearing in mind that hydrogen forms about 35% of the mass of the Sun's central core.

For $T = 1.5 \times 10^7$ K, Eq. 51-3 yields $K_{\text{av}} = 1.9$ keV, and this value is indicated by a vertical line in Fig. 51-9. Note that there are many particles whose energies exceed this mean value.

The curve marked $p(K)$ in Fig. 51-9 is the probability of barrier penetration for two colliding protons. At $K = 6$ keV, for example, we have $p = 2.4 \times 10^{-5}$. This is the probability that two colliding protons, each with $K = 6$ keV, will succeed in penetrating their mutual Coulomb barrier and coming within range of each other's strong nuclear forces. Put another way, on the average, one of every 42,000 such encounters will succeed.

It turns out that the most probable energy for proton–proton fusion events to occur at the Sun's central temperature is about 6 keV. If the energy is much higher, the barrier is more easily penetrated (that is, p is greater), but there are too few protons in the Maxwellian "tail" (n is smaller). If the energy is much lower, there are plenty of protons but the barrier is now too formidable.

SAMPLE PROBLEM 51-5. The deuteron (^{2}H) has a charge $+e$, and its radius has been measured to be 2.1 fm. Two such particles are fired at each other with the same initial kinetic energy K. What must K be if the particles are brought to rest by their mutual Coulomb repulsion when the two deuterons are just "touching"?

Solution Because the two deuterons are momentarily at rest when they "touch" each other, their kinetic energy has all been transformed into electrostatic potential energy associated with the Coulomb repulsion between them. If we treat them as point charges separated by a distance $2R$, we have

$$2K = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r} = \frac{1}{4\pi\epsilon_0}\frac{e^2}{2R},$$

which yields

$$\begin{aligned} K &= \frac{1}{4\pi\epsilon_0}\frac{e^2}{4R} \\ &= \frac{(9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)(1.6 \times 10^{-19}\ \text{C})^2}{4(2.1 \times 10^{-15}\ \text{m})} \\ &= 2.7 \times 10^{-14}\ \text{J} = 170\ \text{keV}. \end{aligned}$$

This quantity provides a reasonable measure of the height of the Coulomb barrier between two deuterons.

51-7 THERMONUCLEAR FUSION IN STARS

Here we consider in more detail the thermonuclear fusion processes that take place in our Sun and in other stars. In the Sun's deep interior, where its mass is concentrated and where most of the energy production takes place, the (central) temperature is 1.5×10^7 K and the central density is on the order of 10^5 kg/m^3, about 13 times the density of lead. The central temperature is so high that, in spite of the high central pressure (2×10^{11} atm), the Sun remains gaseous throughout.

The present composition of the Sun's core is about 35% hydrogen by mass, about 65% helium, and about 1% other elements. At these temperatures the light elements are essentially totally ionized, so that our picture is one of an assembly of protons, electrons, and α particles in random motion.

The Sun radiates at the rate of 3.9×10^{26} W and has been doing so for as long as the solar system has existed, which is about 4.5×10^9 y. It has been known since the 1930s that thermonuclear fusion processes in the Sun's interior account for its prodigious energy output. Before analyzing this further, however, let us dispose of two other possibilities that had been put forward earlier. Consider first chemical reactions such as simple burning. If the Sun, whose mass is 2.0×10^{30} kg, were made of coal and oxygen in just the right proportions for burning, it would last only about 10^3 y, which of course is far too short (see Exercise 41). The Sun, as we shall see, does not burn coal but hydrogen, and in a nuclear furnace, not an atomic or chemical one.

Another possibility is that, as the core of the Sun cools and the pressure there drops, the Sun will shrink under the action of its own strong gravitational forces. By transferring gravitational potential energy to internal energy (just as we do when we drop a stone onto the Earth's surface), the temperature of the Sun's core will rise so that radiation may continue. Calculation shows, however, that the Sun could radiate from this cause for only about 10^8 y, too short by a factor of 25 (see Problem 7).

The Sun's energy is generated by the thermonuclear "burning" (that is, "fusing") of hydrogen to form helium. Figure 51-10 shows the *proton–proton cycle* by which this is accomplished. Note that each reaction shown is a fusion reaction, in that one of the products (^{2}H, ^{3}He, or ^{4}He) has a higher mass number than any of the reacting particles that form it. The reaction energy Q for each reaction shown in Fig. 51-10 is positive. This characterizes an exothermic reaction, with the net release of energy.

The cycle is initiated by the collision of two protons (^{1}H + ^{1}H) to form a deuteron (^{2}H), with the simultaneous creation of a positron (e^+) and a neutrino (ν). The positron very quickly encounters a free electron (e^-) in the Sun and both particles annihilate, their rest energies appearing as two gamma-ray photons (γ). In Fig. 51-10 we follow the consequences of two such events, as indicated in the top row of the figure. Such events are extremely rare. In fact, only once in about 10^{26} proton–proton collisions is a deuteron formed; in the vast majority of cases the colliding protons simply scatter from each other. It is the slowness of this process that regulates the rate of energy production and keeps the Sun from exploding. In spite of this slowness, there are so very many protons in the huge volume of the Sun's core that deuterium is produced there in this way at the rate of about 10^{12} kg/s!

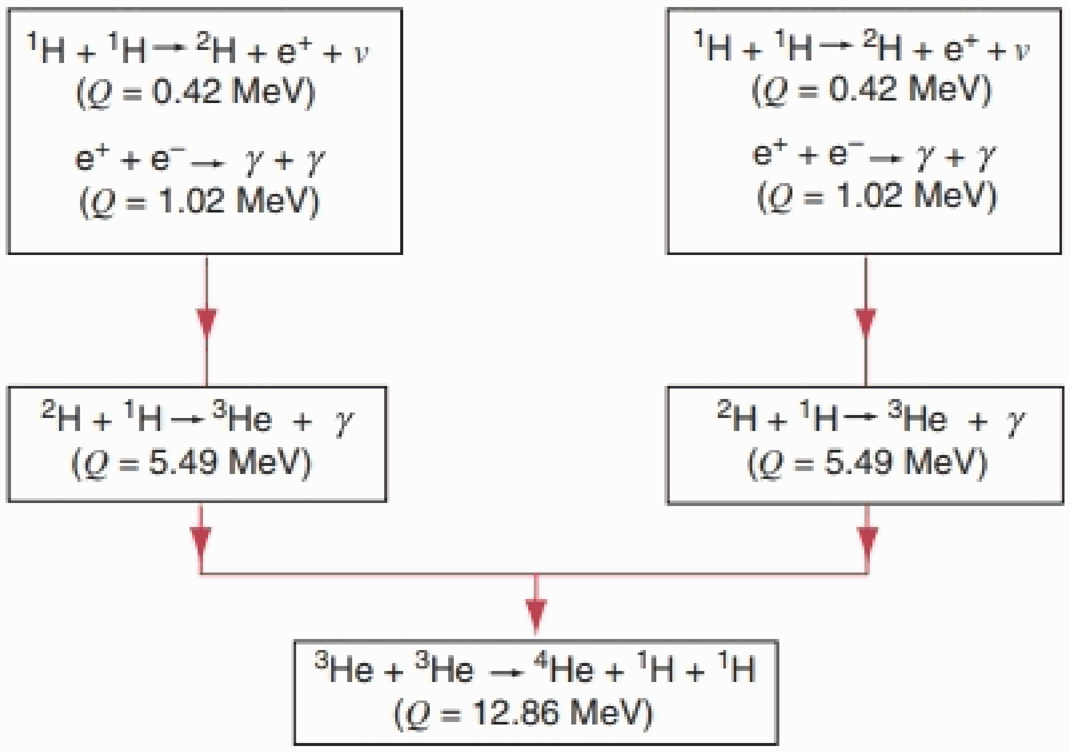

$^1\text{H} + {}^1\text{H} \rightarrow {}^2\text{H} + e^+ + \nu$
(Q = 0.42 MeV)
$e^+ + e^- \rightarrow \gamma + \gamma$
(Q = 1.02 MeV)
$^1\text{H} + {}^1\text{H} \rightarrow {}^2\text{H} + e^+ + \nu$
(Q = 0.42 MeV)
$e^+ + e^- \rightarrow \gamma + \gamma$
(Q = 1.02 MeV)
$^2\text{H} + {}^1\text{H} \rightarrow {}^3\text{He} + \gamma$
(Q = 5.49 MeV)
$^2\text{H} + {}^1\text{H} \rightarrow {}^3\text{He} + \gamma$
(Q = 5.49 MeV)
$^3\text{He} + {}^3\text{He} \rightarrow {}^4\text{He} + {}^1\text{H} + {}^1\text{H}$
(Q = 12.86 MeV)

FIGURE 51-10. The proton–proton cycle that primarily accounts for energy production in the Sun.

Once a deuteron has been produced, it quickly (within a few seconds) collides with another proton and forms a ^{3}He nucleus, as the second row of Fig. 51-10 shows. Two such ^{3}He nuclei may then eventually (within about 10^5 y) collide, forming an α particle (^{4}He) and two protons, as the third row of the figure shows. There are other variations of the proton–proton cycle, involving other light elements, but we concentrate on the principal sequence as represented in Fig. 51-10.

Taking an overall view of the proton–proton cycle, we see that it amounts to the combination of four protons and two electrons to form an α particle, two neutrinos, and six gamma rays:

$$4{}^1\text{H} + 2e^- \rightarrow {}^4\text{He} + 2\nu + 6\gamma. \qquad (51\text{-}4)$$

Now, in a formal way, let us add two electrons to each side of Eq. 51-4, yielding

$$4({}^1\text{H} + e^-) \rightarrow ({}^4\text{He} + 2e^-) + 2\nu + 6\gamma. \qquad (51\text{-}5)$$

The quantities in parentheses then represent *atoms* (not bare nuclei) of hydrogen and of helium.

The energy release in the reaction of Eq. 51-5 is, using the atomic masses of hydrogen and helium,

$$\begin{aligned} Q &= (m_i - m_f)c^2 = [4m({}^1\text{H}) - m({}^4\text{He})]c^2 \\ &= [4(1.007825\text{ u}) - 4.002603\text{ u}](931.5\text{ MeV/u}) \\ &= 26.7\text{ MeV}. \end{aligned}$$

(Gamma-ray photons are massless, and neutrinos have either zero or negligibly small mass; thus neither particle enters into the calculation of the Q value for the fusion reaction.) This same value of Q follows (as it must) by adding up the Q values for the separate steps of the proton–proton cycle in Fig. 51-10.

Not quite all this energy is available as internal energy inside the Sun. About 0.5 MeV is associated with the two neutrinos that are produced in each cycle. Neutrinos are so penetrating that in essentially all cases they escape from the Sun, carrying this energy with them. Some are intercepted by the Earth, bringing us our only direct information about the Sun's interior. Subtracting the neutrino energy leaves 26.2 MeV per cycle available within the Sun. As we show in Sample Problem 51-6, this corresponds to a "heat of combustion" for the nuclear burning of hydrogen into helium of 6.3×10^{14} J/kg of hydrogen consumed. By comparison, the heat of combustion of coal is about 3.3×10^7 J/kg, some 20 million times lower, reflecting roughly the general ratio of energies in nuclear and chemical processes.

We may ask how long the Sun can continue to shine at its present rate before all the hydrogen in its core has been converted into helium. Hydrogen burning has been going on for about 4.5×10^9 y, and calculations show that there is enough available hydrogen left for about 5×10^9 y more. At that time major changes will begin to happen. The Sun's core, which by then will be largely helium, will begin to collapse and to heat up while the outer envelope will expand greatly, perhaps so far as to encompass the Earth's orbit. The Sun will become what astronomers call a *red giant.*

If the core temperature heats up to about 10^8 K, energy can be produced by burning helium to make carbon. Helium does not burn readily, the only possible reaction being

$$^4\text{He} + {}^4\text{He} + {}^4\text{He} \rightarrow {}^{12}\text{C} + \gamma \qquad (Q = +7.3\text{ MeV}).$$

Such a three-body collision of three α particles must occur within 10^{-16} s if the reaction is to go. Nevertheless, if the density and temperature of the helium core are high enough, carbon will be manufactured by the burning of helium in this way.

As a star evolves and becomes still hotter, other elements can be formed by other fusion reactions. However, elements beyond $A = 56$ cannot be manufactured by further fusion processes. The elements with $A = 56$ (^{56}Fe, ^{56}Co, ^{56}Ni) lie near the peak of the binding energy curve of Fig. 50-6, and fusion between nuclides beyond this point involves the consumption, and not the production, of energy. The production of the elements in fusion processes is discussed in Chapter 52.

SAMPLE PROBLEM 51-6. At what rate is hydrogen being consumed in the core of the Sun, assuming that all the radiated energy is generated by the proton–proton cycle of Fig. 51-10?

Solution We have seen that 26.2 MeV appears as internal energy in the Sun for every four protons consumed, a rate of 6.6 MeV/proton. We can express this as

$$\frac{(6.6\text{ MeV/proton})(1.6 \times 10^{-13}\text{ J/MeV})}{1.67 \times 10^{-27}\text{ kg/proton}} = 6.3 \times 10^{14}\text{ J/kg},$$

which tells us that the Sun radiates away 6.3×10^{14} J for every kilogram of protons consumed. The hydrogen consumption rate is then the output power ($= 3.9 \times 10^{26}$ W) divided by this quantity, or

$$\frac{dm}{dt} = \frac{3.9 \times 10^{26}\text{ W}}{6.3 \times 10^{14}\text{ J/kg}} = 6.2 \times 10^{11}\text{ kg/s}.$$

To keep this number in perspective, keep in mind that the Sun's mass is 2.0×10^{30} kg.

51-8 CONTROLLED THERMONUCLEAR FUSION

Thermonuclear reactions have been going on in the universe since its creation in the presumed cosmic "big bang" of some 15 billion years ago. Such reactions have taken place on Earth, however, only since October 1952, when the first fusion (or hydrogen) bomb was exploded. The high temperatures needed to initiate the thermonuclear reaction in this case were provided by a fission bomb used as a trigger.

A sustained and controllable thermonuclear power source—a fusion reactor—is proving much more difficult to achieve. The goal, however, is vigorously pursued because many look to the fusion reactor as the ultimate power source of the future, at least as far as the generation of electricity is concerned.

The proton–proton interaction displayed in Fig. 51-10 is not suitable for use in a terrestrial fusion reactor because the process described in the first row is hopelessly slow. The reaction probability is in fact so small that it cannot be measured in the laboratory. The reaction succeeds under the conditions that prevail in stellar interiors only because of the enormous number of protons available in the high-density stellar cores.

The most attractive reactions for terrestrial use appear to be the deuteron–deuteron (d-d) and the deuteron–triton (d-t) reactions:

$$\text{d-d:}\quad {}^2\text{H} + {}^2\text{H} \rightarrow {}^3\text{He} + \text{n} \qquad (Q = +3.27\ \text{MeV}), \tag{51-6}$$

$$\text{d-d:}\quad {}^2\text{H} + {}^2\text{H} \rightarrow {}^3\text{H} + {}^1\text{H} \qquad (Q = +4.03\ \text{MeV}), \tag{51-7}$$

$$\text{d-t:}\quad {}^2\text{H} + {}^3\text{H} \rightarrow {}^4\text{He} + \text{n} \qquad (Q = +17.59\ \text{MeV}). \tag{51-8}$$

Here *triton* indicates ^{3}H, the nucleus of hydrogen with $A = 3$. Note that each of these reactions is indeed a fusion reaction and has a positive Q value. Deuterium, whose isotopic abundance in normal hydrogen is 0.015%, is available in unlimited quantities as a component of seawater. Tritium (atomic ^{3}H) is radioactive and is not normally found in naturally occurring hydrogen.

There are three basic requirements for the successful operation of a thermonuclear reactor.

1. *A high particle density n.* The number of interacting particles (deuterons, say) per unit volume must be great enough to ensure a sufficiently high deuteron–deuteron collision rate. At the high temperatures required, the deuterium gas would be completely ionized into a neutral *plasma* consisting of deuterons and electrons.

2. *A high plasma temperature T.* The plasma must be hot. Otherwise the colliding deuterons will not be energetic enough to penetrate the mutual Coulomb barrier that tends to keep them apart. In fusion research, temperatures are often reported by giving the corresponding value of kT (not $\frac{3}{2}kT$). A plasma temperature of 43 keV, corresponding to 5.0×10^8 K, has been achieved in the laboratory. This is much higher than the Sun's central temperature (1.3 keV, or 1.5×10^7 K).

3. *A long confinement time τ.* A major problem is containing the hot plasma long enough to ensure that its density and temperature remain sufficiently high. It is clear that no actual solid container can withstand the high temperatures necessarily involved, so special techniques, to be described later, must be employed. By use of one such technique, confinement times greater than 1 s have been achieved.

For the successful operation of a thermonuclear reactor, it can be shown that n, T, and τ must be large enough so

$$n\tau T \geq 50 \times 10^{20}\ \text{keV} \cdot \text{s/m}^3, \tag{51-9}$$

a condition sometimes called *Lawson's criterion.* Equation 51-9 tells us, for a high enough plasma temperature, that we have a choice between confining many particles for a relatively short time or confining fewer particles for a somewhat longer time.

Two techniques have been used to attempt to meet Lawson's criterion and so achieve controlled fusion. *Magnetic confinement* uses magnetic fields to confine the plasma while its temperature is increased. In *inertial confinement,* on the other hand, a small amount of fuel is compressed and heated so rapidly that fusion occurs before the fuel can expand and cool. These techniques are discussed in the following two sections.

Magnetic Confinement

Because a plasma consists of charged particles, its motion can be controlled with magnetic fields. For example, charged particles spiral about the direction of a uniform magnetic field. By suitably varying the field strength, it is possible to design a "magnetic mirror" (see Fig. 32-14) from which particles can be reflected. Another design makes use of toroidal geometry, in which the particles spiral around the axis of a toroid inside a "doughnut-shaped" vacuum chamber. The type of fusion reactor based on this principle, which was first developed in Russia, is called a *tokamak,* which comes from the Russian acronym for "toroidal magnetic chamber." Several large machines of this type have been built and tested.

In a tokamak, there are two components to the magnetic field, as illustrated in Fig. 51-11. The *toroidal* field B_t is the one we usually associate with a toroidal winding of wires; Fig. 51-11 shows one small section of an external coil that contributes to the toroidal field. Because the toroidal field decreases with increasing radius, it is necessary to add a second field component to confine the particles. This *poloidal* component B_p of the field adds to the toroidal component to give the total field a helical structure, as illustrated in Fig. 51-11. The poloidal field is produced by a current i' in the plasma itself, which is induced by a set of

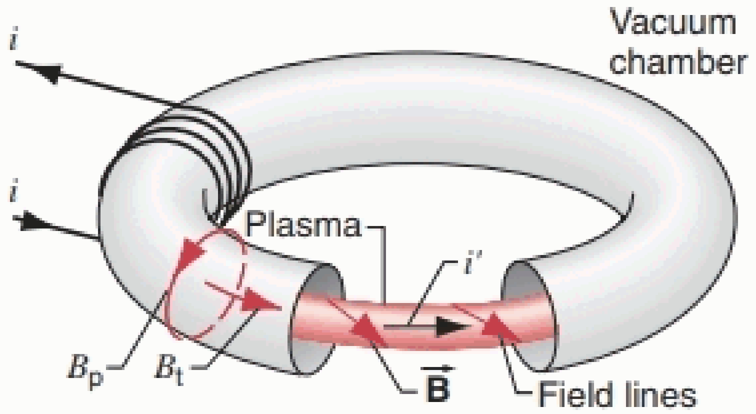


FIGURE 51-11. The toroidal chamber that forms the basis of the tokamak. Note the plasma, the helical magnetic field $\vec{\mathbf{B}}$ that confines it, and the induced current i' that heats it.

windings not illustrated in the figure. This current also serves to heat the plasma. Additional means of heating, such as by firing neutral beams of particles into the plasma, are also necessary to achieve the desired plasma temperature.

Figure 51-12 shows the toroidal vacuum chamber of the Tokamak Fusion Test Reactor at the Princeton Plasma Physics Laboratory. The interior radius of the vacuum chamber is about 2 m, and the major radius of the toroid is 2.5 m. This facility operated until 1998 to develop and test many of the concepts necessary to achieve a workable fusion reactor. It was able to produce 10 MW of fusion power for about 1 second. A larger experimental facility, called the International Thermonuclear Experimental Reactor (ITER) is currently under development as a joint effort by the European Union, Japan, and Russia. This facility, which is planned to be in operation by about the year 2008, is being designed in its initial operation to produce 500 MW of fusion power for times of several minutes.

In designing magnetic confinement devices such as the tokamak, the goal is to increase n, τ, and T. At sufficiently high values of these parameters, Lawson's criterion is satisfied and fusion reactions in the plasma will produce enough energy to equal the energy that must be supplied to heat the plasma. This condition is called "breakeven." At still higher values of these parameters, the device will achieve "ignition," where self-sustaining fusion reactions will occur. Figure 51-13 illustrates the steady progress toward these goals that has been made. Despite the near achievement of the breakeven condition, many formidable engineering problems remain to be solved, and the production of electric power from fusion is likely to be many decades away.

FIGURE 51-12. Workers inside the toroidal chamber of the Tokamak Fusion Test Reactor at Princeton University.

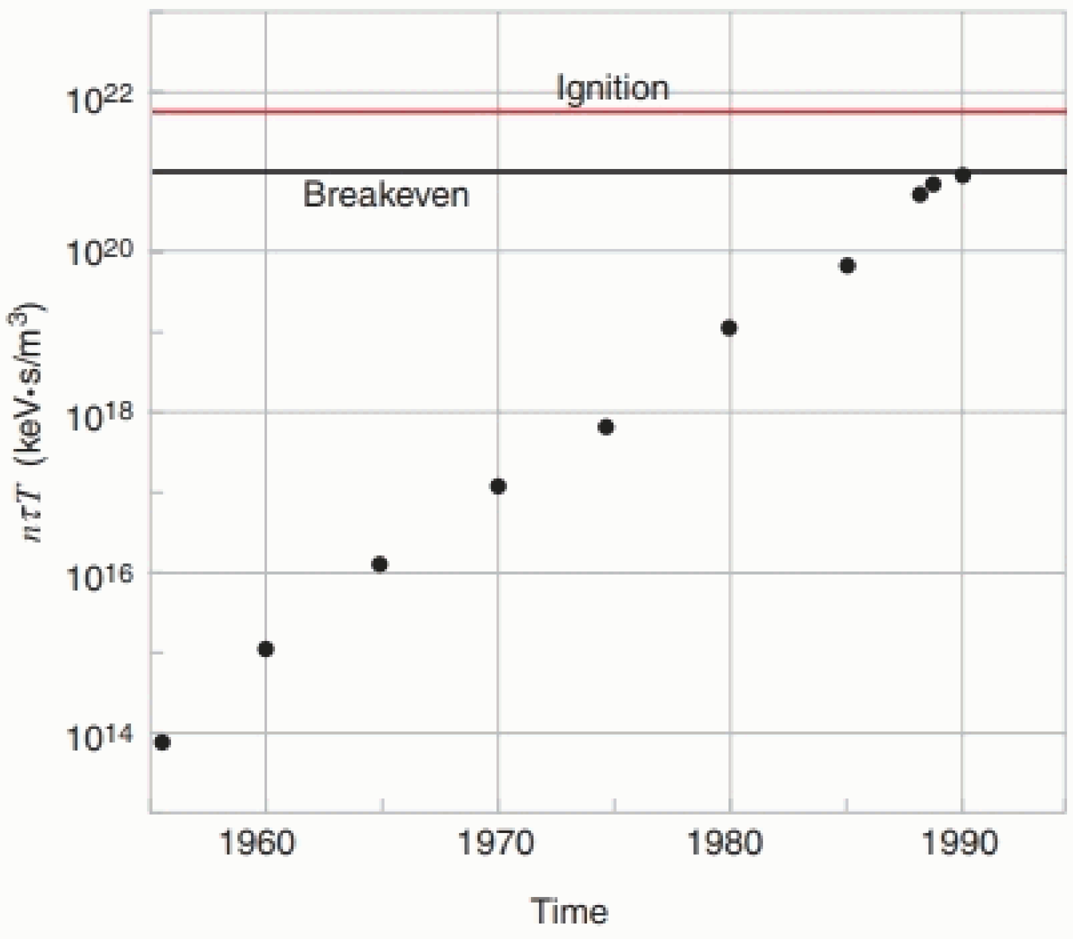


FIGURE 51-13. The approach to breakeven and ignition in controlled fusion reactors, shown as a plot of $n\tau T$ against time.

SAMPLE PROBLEM 51-7. The Tokamak Fusion Test Reactor at Princeton has achieved a confinement time of 400 ms. (*a*) If the plasma temperature, measured in energy units as kT, is 20 keV, what must be the density of particles in the plasma if Lawson's criterion for ignition is to be satisfied? (*b*) How does this number compare with the particle density of the atoms of an ideal gas at standard conditions? (*c*) Under these same conditions of density and temperature, what confinement time is necessary to achieve the breakeven condition?

Solution (*a*) Using Lawson's criterion (Eq. 51-9), we must have

$$\begin{aligned} n &= \frac{50 \times 10^{20}\ \text{keV} \cdot \text{s/m}^3}{\tau T} \\ &= \frac{50 \times 10^{20}\ \text{keV} \cdot \text{s/m}^3}{(400 \times 10^{-3}\ \text{s})(20\ \text{keV})} = 6.3 \times 10^{20}\ \text{m}^{-3}. \end{aligned}$$

(*b*) The number density of atoms in an ideal gas at standard conditions is given by $n' = N_A/V_m$, where N_A is the Avogadro constant and V_m $(= 2.24 \times 10^{-2}\ \text{m}^3/\text{mol})$ is the molar volume of an ideal gas at standard conditions, which gives

$$n' = \frac{N_A}{V_m} = \frac{6.02 \times 10^{23}\ \text{mol}^{-1}}{2.24 \times 10^{-2}\ \text{m}^3/\text{mol}} = 2.7 \times 10^{25}\ \text{m}^{-3}.$$

The particle density of the plasma we found in part (*a*) is smaller than that of an ideal gas by a factor of about 4×10^4.

(c) Figure 51-13 shows that for breakeven we must have $n\tau T \approx 1 \times 10^{21}$ keV · s/m^3, or about $\frac{1}{5}$ of Lawson's criterion for ignition. With the same density and temperature as part (a), we must therefore have about $\frac{1}{5}$ of the confinement time, or about 80 ms.

Inertial Confinement

A second technique for confining plasma so that thermonuclear fusion can take place is called *inertial confinement.* In terms of Lawson's criterion (Eq. 51-9), it involves working with extremely high particle densities n for extremely short confinement times τ. These times are arranged to be so short that the fusion episode is over before the particles of the plasma have time to move appreciably from the positions they occupy at the onset of fusion. The interacting particles are confined by their own inertia.

Laser fusion, which relies on the inertial-confinement principle, is being investigated in laboratories throughout the world. At the Lawrence Livermore National Laboratory, for example, in the NOVA laser fusion project (see Fig. 51-14) deuterium–tritium fuel pellets, each smaller than a grain of sand (see Fig. 51-15), are to be "zapped" by 10 synchronized, high-powered laser pulses, symmetrically arranged around the pellet. The laser pulses are designed to deliver in total some 35 kJ of energy to each fuel pellet in less than a nanosecond. This is a delivered power of 4×10^{13} W during the pulse, which is roughly 100 times the total installed electric power-generating capacity of the world!

The laser pulse energy serves to heat the fuel pellet, ionizing it to a plasma and—it is hoped—raising its temperature to around 10^8 K. As the surface layers of the pellet evaporate at these high thermal speeds, the reaction force of the escaping particles compresses the core of the pellet, increasing its density by a factor of perhaps 10^3. If all these things happened, then conditions would be right for thermonuclear fusion to occur in the core of the highly compressed pellet of plasma, the fusion reaction being the d-t reaction given in Eq. 51-8.

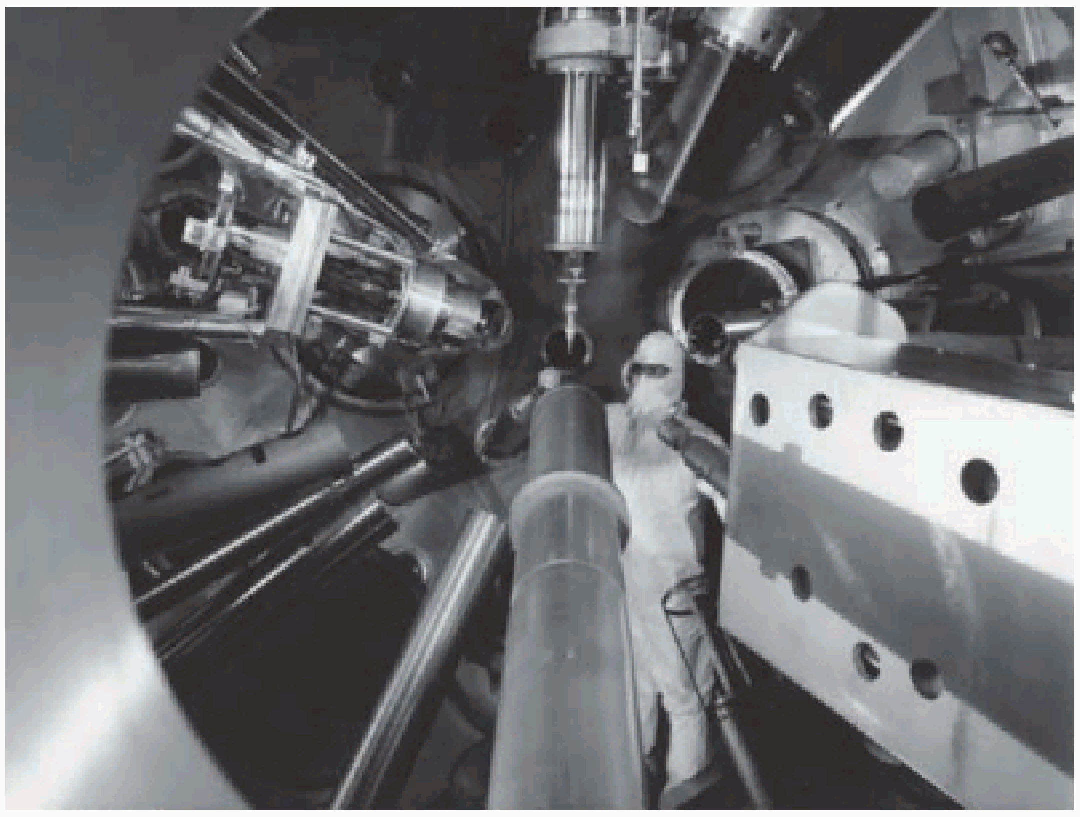


FIGURE 51-14. The target chamber of the NOVA inertial confinement fusion facility at the Lawrence Livermore National Laboratory. The photo shows several of the 10 laser beam tubes.

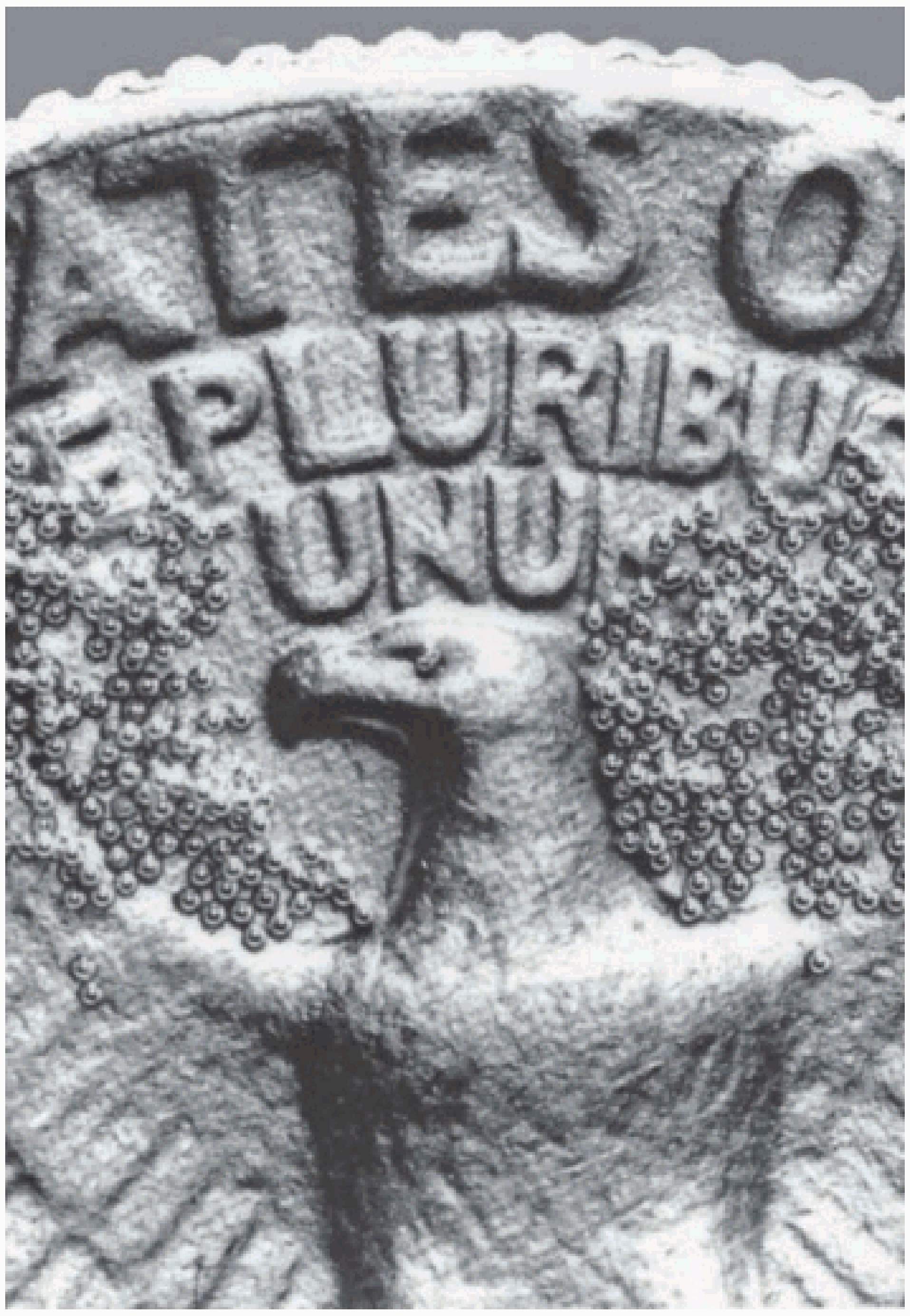


FIGURE 51-15. The tiny spheres, shown resting on a dime, are deuterium–tritium fuel pellets for use in inertial confinement experiments.

In an operating thermonuclear reactor of the laser fusion type, it is visualized that fuel pellets would be exploded, like miniature hydrogen bombs, at the rate of perhaps 10–100 per second. The energetic emerging particles of the fusion reaction (^{4}He and n) might be absorbed in a "blanket" consisting of a moving stream of molten lithium, heating it up. Internal energy would then be extracted from the lithium stream at another location and used to generate steam, just as in a fission reactor or a fossil-fuel power plant. Lithium would be a suitable choice for a heat-transfer medium because the energetic neutron would, with high probability, deliver up its energy to the "blanket" by the reaction

$$^6\text{Li} + \text{n} \rightarrow {}^4\text{He} + {}^3\text{H}.$$

The two charged particles would readily be brought to rest in the lithium. The tritium produced in the reaction can be extracted for use as fuel in the reactor.

A new inertial confinement fusion experimental facility is under construction at the Livermore laboratory. The National Ignition Facility, expected to be completed in 2002, will bring 192 laser beams onto a deuterium–tritium target, delivering an energy of 1.8 MJ. Like magnetic confinement, inertial confinement fusion remains a subject of active research and development.

SAMPLE PROBLEM 51-8. Suppose that a fuel pellet in a laser fusion device is made of a liquid deuterium–tritium mixture containing equal numbers of deuterium and tritium atoms. The density d ($=200\ \mathrm{kg/m^3}$) of the pellet is increased by a factor of 10^3 by the action of the laser pulses. (*a*) How many particles per unit volume (either deuterons or tritons) does the pellet contain in its compressed state? (*b*) At a plasma temperature corresponding to $kT = 50$ keV, for how long must this particle density be maintained to satisfy Lawson's criterion?

Solution (*a*) We can write, for the density d' of the compressed pellet,

$$d' = 10^3 d = m_\mathrm{d}\frac{n}{2} + m_\mathrm{t}\frac{n}{2},$$

in which n is the number of particles per unit volume (either deuterons or tritons) in the compressed pellet, m_d is the mass of a deuterium atom, and m_t is the mass of a tritium atom. These atomic masses are related to the Avogadro constant N_A and to the corresponding molar masses (M_d and M_t) by

$$m_\mathrm{d} = M_\mathrm{d}/N_\mathrm{A} \quad \text{and} \quad m_\mathrm{t} = M_\mathrm{t}/N_\mathrm{A}.$$

Combining these equations and solving for n lead to

$$\begin{aligned} n &= \frac{2d'N_\mathrm{A}}{M_\mathrm{d} + M_\mathrm{t}} \\ &= \frac{(2)(10^3 \times 200\ \mathrm{kg/m^3})(6.02 \times 10^{23}\ \mathrm{mol^{-1}})}{2.0 \times 10^{-3}\ \mathrm{kg/mol} + 3.0 \times 10^{-3}\ \mathrm{kg/mol}} \\ &= 4.8 \times 10^{31}\ \mathrm{m^{-3}}. \end{aligned}$$

(*b*) From Lawson's criterion (Eq. 51-9), we have

$$\begin{aligned} \tau &= \frac{50 \times 10^{20}\ \mathrm{keV \cdot s/m^{-3}}}{nT} \\ &= \frac{50 \times 10^{20}\ \mathrm{keV \cdot s/m^{-3}}}{(4.8 \times 10^{31}\ \mathrm{m^{-3}})(50\ \mathrm{keV})} = 2 \times 10^{-12}\ \mathrm{s}. \end{aligned}$$

The pellet must remain compressed for at least this long if breakeven operation is to occur.

A comparison with Sample Problem 51-7 shows that, unlike tokamak operation, laser fusion seeks to operate in the realm of very high particle densities and correspondingly very short confinement times.

MULTIPLE CHOICE

51-1 The Atom and the Nucleus

51-2 Nuclear Fission: The Basic Process

1. Consider the fission reaction $^{235}\mathrm{U} + \mathrm{n} \rightarrow {}^{99}\mathrm{X} + {}^{134}\mathrm{Te} + x\mathrm{n}$.
   (*a*) What chemical element does X represent?
   (A) Sr (B) Zr (C) Nb
   (D) The answer cannot be determined without more information.
   (*b*) What is the value of x?
   (A) 1 (B) 2 (C) 2.47 (D) 3
2. Why are the fission fragments usually radioactive?
   (A) They come originally from radioactive $^{235}\mathrm{U}$.
   (B) They have a large neutron excess.
   (C) They have a large binding energy per nucleon.
   (D) They are moving at high speed.
3. In a particular fission process of a uranium atom at rest, fragments with mass numbers A_1 and A_2 are produced. A good estimate of the ratio of kinetic energies K_1/K_2 is
   (A) A_2/A_1. (B) $(A_2/A_1)^2$. (C) A_1/A_2. (D) $(A_1/A_2)^2$.

51-3 Theory of Nuclear Fission

51-4 Nuclear Reactors: The Basic Principles

4. In a nuclear reactor, the function of the moderator is
   (A) to absorb neutrons.
   (B) to keep the reactor from going critical.
   (C) to slow down the neutrons.
   (D) to absorb heat from the core.

51-5 A Natural Reactor

51-6 Thermonuclear Fusion: The Basic Process

5. Extremely high-speed protons are not important in thermonuclear fusion because
   (A) they have more difficulty penetrating the Coulomb barrier.
   (B) they move too fast to collide with another proton.
   (C) there is a relativistic increase in the proton's electric charge.
   (D) there are relatively very few of them.

51-7 Thermonuclear Fusion in Stars

6. On the average, the universe is about 25% helium and 75% hydrogen. The Sun's core is about 65% helium. This excess helium is
   (A) needed for the fusion reactions that produce the Sun's energy.
   (B) only a random fluctuation in the helium concentration of the universe.
   (C) a product of the fusion reactions in the Sun.
   (D) a catalyst that enhances the production of energy by fusion.

51-8 Controlled Thermonuclear Fusion

7. What is the main difficulty associated with the fusion process as a source of electrical power?
   (A) The scarcity of fuel
   (B) The Coulomb barrier
   (C) The radioactivity of the products
   (D) The danger of an explosion
8. Carbon and oxygen are also abundant and cheap elements. Why do we not try to use them instead of hydrogen as fuel for a fusion reactor?
   (A) The energy obtained per nucleon is much smaller.
   (B) The Coulomb barrier is much higher.
   (C) Both (A) and (B).
   (D) Neither (A) nor (B).

9. Compared with the d-d reactions, the energy release in the d-t reaction is so large because
(A) tritium (^3H) is radioactive and therefore packs more energy.
(B) the reaction product ^4He is very tightly bound.
(C) more nucleons participate in the d-t reaction.
(D) the Coulomb barrier is lower.

QUESTIONS

1. If it is so much harder to get a nucleon out of a nucleus than to get an electron out of an atom, why try?

2. Can you say, from examining Table 51-1, that one source of energy, or of power, is better than another? If not, what other considerations enter?

3. To which of the processes in Table 51-1 does the relationship $E = \Delta m\, c^2$ apply?

4. Of the two fission fragment tracks shown in Fig. 51-1, which fragment has the larger (*a*) momentum, (*b*) kinetic energy, (*c*) speed, (*d*) mass?

5. In the generalized equation for the fission of ^{235}U by thermal neutrons, $^{235}U + n \rightarrow X + Y + bn$, do you expect the Q of the reaction to depend on the identity of X and Y?

6. Is the fission fragment curve of Fig. 51-2 necessarily symmetrical about its central minimum? Explain your answer.

7. In the chain decays of the primary fission fragments (see Eq. 51-2) why do no β^+ decays occur?

8. The half-life of ^{235}U is 7.0×10^8 y. Discuss the assertion that if it had turned out to be shorter by a factor of 10 or so, there would not be any atomic bombs today.

9. ^{238}U is not fissionable by thermal neutrons. What minimum neutron energy do you think would be necessary to induce fission in this nuclide?

10. The half-life for the decay of ^{235}U by alpha emission is 7×10^8 y; by spontaneous fission, acting alone, it would be 3×10^{17} y. Both are barrier-tunneling processes, as Fig. 50-9 and Fig. 51-4 reveal. Why this enormous difference in barrier-tunneling probability?

11. Compare fission with alpha decay in as many ways as possible. How can a thermal neutron deliver several million electron-volts of excitation energy to a nucleus that absorbs it, as in Fig. 51-3*a*? The neutron has essentially no energy to start with!

12. The binding energy curve of Fig. 50-6 tells us that any nucleus more massive than $A \approx 56$ can release energy by the fission process. Only very massive nuclides seem to do so, however. Why cannot lead, for example, release energy by the fission process?

13. By bombardment of heavy nuclides in the laboratory it is possible to prepare other heavy nuclides that decay, at least in part, by *spontaneous fission*. That is, after a certain mean life they spontaneously break up into two major fragments. Can you explain this on the basis of the theory of Bohr and Wheeler?

14. Slow neutrons are more effective than fast ones in inducing fission. Can you make that plausible? (*Hint:* Consider how the de Broglie wavelength of a neutron might be related to its capture cross section in ^{235}U.)

15. Compare a nuclear reactor with a coal fire. In what sense does a chain reaction occur in each? What is the energy-releasing mechanism in each case?

16. Not all neutrons produced in a reactor are destined to initiate a fission event. What happens to those that do not?

17. Explain what is meant by the statement that in a reactor core neutron leakage is a surface effect and neutron production is a volume effect.

18. Explain the purpose of the moderator in a nuclear reactor. Is it possible to design a reactor that does not need a moderator? If so, what are some of the advantages and disadvantages of such a reactor?

19. Describe how to operate the control rods of a nuclear reactor (*a*) during initial start-up; (*b*) to reduce the power level; and (*c*) on a long-term basis, as fuel is consumed.

20. A reactor is operating at full power with its multiplication factor k adjusted to unity. If the reactor is now adjusted to operate stably at half power, what value must k now assume?

21. Separation of the two isotopes ^{238}U and ^{235}U from natural uranium requires a physical method, such as diffusion, rather than a chemical method. Explain why.

22. A piece of pure ^{235}U (or ^{239}Pu) will spontaneously explode if it is larger than a certain "critical size." A smaller piece will not explode. Explain.

23. What can you say, if anything, about the value of the multiplication factor k in an atomic (fission) bomb?

24. The Earth's core is thought to be mostly iron because, during the formation of the Earth, heavy elements such as iron would have sunk toward the Earth's center and lighter elements, such as silicon, would have floated upward to form the Earth's crust. However, iron is far from the heaviest element. Why is the Earth's core not made of uranium?

25. From information given in the text, collect and write down the approximate heights of the Coulomb barriers for (*a*) the alpha decay of ^{238}U, (*b*) the fission of ^{235}U by thermal neutrons, and (*c*) the head-on collision of two deuterons.

26. The Sun's energy is assumed to be generated by nuclear reactions such as the proton–proton cycle. What alternative ways of generating solar energy were proposed in the past, and why were they rejected?

27. Elements up to mass number ≈ 56 are created by thermonuclear fusion in the cores of stars. Why are heavier elements not also created by this process?

28. Do you think that the thermonuclear fusion reaction controlled by the two curves plotted in Fig. 51-9 necessarily has its maximum effectiveness for the energy at which the two curves cross each other? Explain your answer.

29. In Fig. 51-9, are you surprised that, as judged by the areas under the curve marked $n(K)$, the number of particles with $K > K_{\text{av}}$ is smaller than the number with $K < K_{\text{av}}$, where K_{av} is the average thermal energy?

30. The uranium nuclides present in the Earth today were originally built up and spewed into space during the explosion of stars, so-called supernova events. These explosions, which occurred before the formation of our solar system, represent the collapse of stars under their own gravity. Can you then say that the energy derived from fission was once stored in a gravitational field? Does fission energy then, in this limited sense, have something in common with energy derived from hydroelectric sources?

31. Why does it take so long ($\sim 10^6$ y!) for gamma-ray photons generated by nuclear reactions in the Sun's central core to diffuse to the surface? What kinds of interactions do they have with the protons, α particles, and electrons that make up the core?

32. The primordial matter of the early universe is thought to have been largely hydrogen. Where did all the silicon in the Earth come from? All the gold?

33. Do conditions at the core of the Sun satisfy Lawson's criterion for a sustained thermonuclear fusion reaction? Explain.

34. To achieve ignition in a tokamak, why do you need a high plasma temperature? A high density of plasma particles? A long confinement time?

35. Which would generate more radioactive waste products, a fission reactor or a fusion reactor?

36. Does Lawson's criterion hold both for tokamaks and for laser fusion devices?

EXERCISES

51-1 The Atom and the Nucleus

51-2 Nuclear Fission: The Basic Process

1. You wish to produce 1.0 GJ of energy. Calculate and compare (*a*) the amount of coal needed if you obtain the energy by burning coal and (*b*) the amount of natural uranium needed if you obtain the energy by fission in a reactor. Assume that the combustion of 1.0 kg of coal releases 2.9×10^7 J; the fission of 1.0 kg of uranium in a reactor releases 8.2×10^{13} J.

2. In the United States, coal commonly contains about 3 parts per million (3 ppm) of fissionable uranium and thorium. Calculate and compare (*a*) the energy derived from burning 100 kg of coal and (*b*) the energy that could be derived from the fission of the fissionable impurities that remain in its ashes. Assume that the combustion of 1 kg of coal releases 2.9×10^7 J; the fission of 1 kg of uranium or thorium in a reactor releases 8.2×10^{13} J.

3. (*a*) How many atoms are contained in 1.00 kg of pure ^{235}U? (*b*) How much energy, in joules, is produced by the complete fissioning of 1.00 kg of ^{235}U? Assume $Q = 200$ MeV. (*c*) For how many years would this energy light a 100-W lamp?

4. At what rate must ^{235}U nuclei undergo fission by neutrons to generate 2.00 W? Assume that $Q = 200$ MeV.

5. Verify that, as reported in Table 51-1, the fission of 1 kg of ^{235}U could keep a 100-W lamp burning for 3×10^4 y.

6. The fission properties of the plutonium isotope ^{239}Pu are very similar to those of ^{235}U. The average energy released per fission is 180 MeV. How much energy, in joules, is liberated if all the atoms in 1.00 kg of pure ^{239}Pu undergo fission?

7. Very occasionally a ^{233}U nucleus, having absorbed a neutron, breaks up into *three* fragments. If two of these fragments are identified chemically as isotopes of chromium and gallium and if no prompt neutrons are involved, what is at least one possibility for the identity of the fragments? Consult a nuclidic chart or table.

8. Show that, in Sample Problem 51-1, there is no need to take the masses of the electrons emitted during the beta decay of the primary fission fragments explicitly into account.

9. ^{235}U decays by alpha emission with a half-life of 7.04×10^8 y. It also decays (rarely) by spontaneous fission, and if the alpha decay did not occur, its half-life due to this process alone would be 3.50×10^{17} y. (*a*) At what rate do spontaneous fission decays occur in 1.00 g of ^{235}U? (*b*) How many alpha-decay events occur for every spontaneous fission event?

51-3 Theory of Nuclear Fission

10. Fill in the following table, which refers to the generalized fission reaction

$$^{235}\text{U} + \text{n} \rightarrow X + Y + b\text{n}.$$

X	Y	b
^{140}Xe	—	1
^{139}I	—	2
—	^{100}Zr	2
^{141}Cs	^{92}Rb	—

11. Calculate the distintegration energy Q for the spontaneous fission of ^{52}Cr into two equal fragments. The needed masses are ^{52}Cr, 51.94012 u; and ^{26}Mg, 25.982593 u. Discuss your result.

12. Calculate the disintegration energy Q for the fission of ^{98}Mo into two equal parts. The needed masses are ^{98}Mo, 97.905408 u; and ^{49}Sc, 48.950024 u. If Q turns out to be positive, discuss why this process does not occur spontaneously.

13. Calculate the energy released in the fission reaction

$$^{235}\text{U} + \text{n} \rightarrow {}^{141}\text{Cs} + {}^{92}\text{Rb} + 3\text{n}.$$

Needed atomic masses are

^{235}U	235.043923 u	^{92}Rb	91.919726 u
^{141}Cs	140.920044 u	n	1.008665 u.

14. ^{238}Np has a barrier energy for fission of 4.2 MeV. To remove a neutron from this nuclide requires an energy expenditure of 5.0 MeV. Is ^{237}Np fissionable by thermal neutrons?

15. Consider the fission of ^{238}U by fast neutrons. In one fission event no neutrons were emitted and the final stable end products, after the beta decay of the primary fission fragments, were ^{140}Ce and ^{99}Ru. (*a*) How many beta-decay events were there in the two beta-decay chains, considered together? (*b*) Calculate Q. The relevant atomic masses are

^{238}U	238.050783 u	^{140}Ce	139.905434 u
n	1.008665 u	^{99}Ru	98.905939 u.

16. In a particular fission event of ^{235}U by slow neutrons, it happens that no neutron is emitted and that one of the primary fission fragments is ^{83}Ge. (*a*) What is the other fragment? (*b*) How is the disintegration energy $Q = 170$ MeV split between the two fragments? (*c*) Calculate the initial speed of each fragment.

51-4 Nuclear Reactors: The Basic Principles

17. Many fear that helping additional nations develop nuclear power reactor technology will increase the likelihood of nuclear war because reactors can be used not only to produce energy but, as a by-product through neutron capture with inexpensive ^{238}U, to make ^{239}Pu, which is a "fuel" for nuclear bombs (*breeder* reactors). What simple series of reactions involving neutron capture and beta decay would yield this plutonium isotope?

18. A 190-MW fission reactor consumes half its fuel in 3 years. How much ^{235}U did it contain initially? Assume that all the energy generated arises from the fission of ^{235}U and that this nuclide is consumed only by the fission process. See Sample Problem 51-3.

19. Repeat Exercise 18 taking into account nonfission neutron capture by the ^{235}U. See Sample Problem 51-3.

20. Among the many fission products that may be extracted chemically from the spent fuel of a nuclear power reactor is ^{90}Sr ($t_{1/2} = 29$ y). It is produced in typical large reactors at the rate of about 18 kg/y. By its radioactivity it generates thermal energy at the rate of 2.3 W/g. (*a*) Calculate the effective disintegration energy Q_{eff} associated with the decay of a ^{90}Sr nucleus. (Q_{eff} includes contributions from the decay of the ^{90}Sr daughter products in its decay chain but not from neutrinos, which escape totally from the sample.) (*b*) It is desired to construct a power source generating 150 W (electric) to use in operating electronic equipment in an underwater acoustic beacon. If the source is based on the thermal energy generated by ^{90}Sr and if the efficiency of the thermal–electric conversion process is 5.0%, how much ^{90}Sr is needed? The atomic mass of ^{90}Sr is 89.9 u.

21. The neutron generation time t_{gen} in a reactor is the average time between one fission and the fissions induced by the neutrons emitted in that fission. Suppose that the power output of a reactor at time $t = 0$ is P_0. Show that the power output a time t later is $P(t)$, where

$$P(t) = P_0\, k^{t/t_{\text{gen}}},$$

where k is the multiplication factor. Note that for constant power output, $k = 1$.

22. The neutron generation time (see Exercise 21) of a particular power reactor is 1.3 ms. It is generating energy at the rate of 1200 MW. To perform certain maintenance checks, the power level must be temporarily reduced to 350 MW. It is desired that the transition to the reduced power level take 2.6 s. To what (constant) value should the multiplication factor be set to effect the transition in the desired time?

23. The neutron generation time t_{gen} (see Exercise 21) in a particular reactor is 1.0 ms. If the reactor is operating at a power level of 500 MW, about how many free neutrons (neutrons that will subsequently induce a fission) are present in the reactor at any moment?

24. A reactor operates at 400 MW with a neutron generation time of 30 ms. If its power increases for 5.0 min with a multiplication factor of 1.0003, find the power output at the end of the 5.0 min. See Exercise 21.

25. The thermal energy generated when radiations from radionuclides are absorbed in matter can be used as the basis for a small power source for use in satellites, remote weather stations, and so on. Such radionuclides are manufactured in abundance in nuclear power reactors and may be separated chemically from the spent fuel. One suitable radionuclide is ^{238}Pu ($t_{1/2} = 87.7$ y), which is an alpha emitter with $Q = 5.59$ MeV. At what rate is thermal energy generated in 1.00 kg of this material?

26. One possible method for revealing the presence of concealed nuclear weapons is to detect the neutrons emitted in the spontaneous fission of ^{240}Pu in the warhead. In an actual trial, a neutron detector of area 2.5 m^2, carried on a helicopter, measured a neutron flux of 4.0 s^{-1} at a distance of 35 m from a missile warhead. Estimate the mass of ^{240}Pu in the warhead. The half-life for spontaneous fission in ^{240}Pu is 1.34×10^{11} y, and 2.5 neutrons, on the average, are emitted in each fission.

51-5 A Natural Reactor

27. How far back in time would natural uranium have been a practical reactor fuel, with a ^{235}U/^{238}U ratio of 3.00%? See Sample Problem 51-4.

28. The natural fission reactor discussed in Section 51-5 is estimated to have generated 15 gigawatt-years of energy during its lifetime. (*a*) If the reactor lasted for 200,000 y, at what average power level did it operate? (*b*) How much ^{235}U did it consume during its lifetime?

29. Some uranium samples from the natural reactor site described in Section 51-5 were found to be slightly *enriched* in ^{235}U rather than depleted. Account for this in terms of neutron absorption by the abundant isotope ^{238}U and the subsequent beta and alpha decay of its products.

51-6 Thermonuclear Fusion: The Basic Process

30. For how long could the fusion of 1.00 kg of deuterium by the reaction

$$^2\text{H} + {}^2\text{H} \rightarrow {}^3\text{He} + \text{n} \qquad (Q = +3.27\ \text{MeV})$$

keep a 100-W lamp burning? The atomic mass of deuterium is 2.014 u.

31. Calculate the height of the Coulomb barrier for the head-on collision of two protons. The effective radius of a proton may be taken to be 0.80 fm. See Sample Problem 51-5.

32. The equation of the curve $n(K)$ in Fig. 51-9 is

$$n(K) = \frac{2N}{\sqrt{\pi}} \frac{K^{1/2}}{(kT)^{3/2}}\, e^{-K/kT},$$

where N is the total density of particles. At the center of the Sun the temperature is 1.5×10^7 K and the mean proton energy K_{av} is 1.9 keV. Find the ratio of the density of protons at 5.0 keV to that at the mean proton energy.

33. Methods other than heating the material have been suggested for overcoming the Coulomb barrier for fusion. For example, one might consider using particle accelerators. If you were to use two of them to accelerate two beams of deuterons directly toward each other so as to collide "head-on," (*a*) what voltage would each require to overcome the Coulomb barrier? (*b*) Would this voltage be difficult to achieve? (*c*) Why do you suppose this method is not presently used?

51-7 Thermonuclear Fusion in Stars

34. We have seen that Q for the overall proton–proton cycle is 26.7 MeV. How can you relate this number to the Q values for the three reactions that make up this cycle, as displayed in Fig. 51-10?

35. Show that the energy released when three alpha particles fuse to form ^{12}C is 7.27 MeV. The atomic mass of ^{4}He is 4.002603 u and of ^{12}C is 12.000000 u.

36. At the central core of the Sun the density is 1.5×10^5 kg/m^3 and the composition is essentially 35% hydrogen by mass and 65% helium. (*a*) What is the density of protons at the Sun's core? (*b*) What is the ratio of this to the density of particles for an ideal gas at standard conditions of temperature and pressure?

37. Calculate and compare the energy in MeV released by (*a*) the fusion of 1.0 kg of hydrogen deep within the Sun and (*b*) the fission of 1.0 kg of ^{235}U in a fission reactor.

38. The Sun has a mass of 2.0×10^{30} kg and radiates energy at the rate of 3.9×10^{26} W. (*a*) At what rate does the mass of the Sun decrease? (*b*) What fraction of its original mass has the Sun lost in this way since it began to burn hydrogen, about 4.5×10^9 y ago?

39. Let us assume that the core of the Sun has one-eighth the Sun's mass and is compressed within a sphere whose radius is one-fourth of the solar radius. We assume further that the composition of the core is 35% hydrogen by mass and that essentially all of the Sun's energy is generated there. If the Sun continues to burn hydrogen at the rate calculated in Sample Problem 51-6, how long will it be before the hydrogen is entirely consumed? The Sun's mass is 2.0×10^{30} kg.

40. Verify the Q values reported for the reactions in Fig. 51-10. The needed atomic masses are

^{1}H	1.007825 u	^{3}He	3.016029 u
^{2}H	2.014102 u	^{4}He	4.002603 u
$e^{\pm}$	0.0005486 u.		

(*Hint:* Distinguish carefully between atomic and nuclear masses, and take the positrons properly into account.)

41. Coal burns according to

$$C + O_2 \rightarrow CO_2.$$

The heat of combustion is 3.3×10^7 J/kg of atomic carbon consumed. (*a*) Express this in terms of energy per carbon atom. (*b*) Express it in terms of energy per kilogram of the initial reactants, carbon and oxygen. (*c*) Suppose that the Sun (mass $= 2.0 \times 10^{30}$ kg) were made of carbon and oxygen in combustible proportions and that it continued to radiate energy at its present rate of 3.9×10^{26} W. How long would it last?

42. After converting all its hydrogen to helium, a particular star is 100% helium in composition. It now proceeds to convert the helium to carbon via the triple-alpha process

$$^4\text{He} + {}^4\text{He} + {}^4\text{He} \rightarrow {}^{12}\text{C} + \gamma \qquad (Q = 7.27\ \text{MeV}).$$

The mass of the star is 4.6×10^{32} kg, and it generates energy at the rate of 5.3×10^{30} W. How long will it take to convert all the helium to carbon?

51-8 Controlled Thermonuclear Fusion

43. Verify the Q values reported in Eqs. 51-6, 51-7, and 51-8. The needed masses are

^{1}H	1.007825 u	^{3}He	3.016029 u
^{2}H	2.014102 u	^{4}He	4.002603 u
^{3}H	3.016049 u	n	1.008665 u.

44. Ordinary water consists of roughly 0.015% by mass of "heavy water," in which one of the two hydrogens is replaced with deuterium, ^{2}H. How much average fusion power could be obtained if we "burned" all the ^{2}H in 1 liter of water in 1 day through the reaction $^2\text{H} + {}^2\text{H} \rightarrow {}^3\text{He} + \text{n}$ ($Q = 3.27$ MeV)?

45. In the deuteron–triton fusion reaction of Eq. 51-8, how is the reaction energy Q shared between the α particle and the neutron (that is, calculate the kinetic energies K_α and K_n)? Neglect the relatively small kinetic energies of the two combining particles.

46. Figure 51-16 shows an idealized representation of a hydrogen bomb. The fusion fuel is lithium deuteride (LiD). The high temperature, particle density, and neutrons to induce fusion are provided by an atomic (fission) bomb "trigger." The fusion reactions are

$$^6\text{Li} + \text{n} \rightarrow {}^3\text{H} + {}^4\text{He}$$

and

$$^2\text{H} + {}^3\text{H} \rightarrow {}^4\text{He} + \text{n} \qquad (Q = 17.59\ \text{MeV}),$$

the tritium (^{3}H) produced in the first reaction fusing with the deuterium (D) in the fuel; see Eq. 51-8. By calculating Q for the first reaction, find the mass of LiD required to produce a fusion yield of 1 megaton of TNT ($= 2.6 \times 10^{28}$ MeV). Needed atomic masses are

^{6}Li	6.015122 u	^{4}He	4.002603 u
^{3}H	3.016049 u	n	1.008665 u.

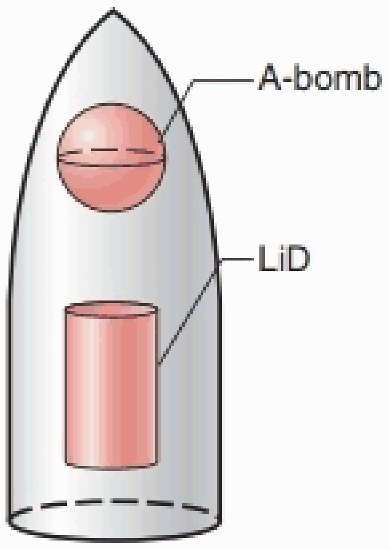


FIGURE 51-16. Exercise 46.

47. Assume that a plasma temperature of 1.3×10^8 K is reached in a laser-fusion device. (*a*) What is the most probable speed of a deuteron at this temperature? (*b*) How far would such a deuteron move in the confinement time calculated in Sample Problem 51-8?

PROBLEMS

1. Assume that immediately after the fission of ^{236}U according to Eq. 51-2, the resulting ^{140}Xe and ^{94}Sr nuclei are just touching at their surfaces. (*a*) Assuming the nuclei to be spherical, calculate the Coulomb potential energy (in MeV) of repulsion between the two fragments. (*Hint:* Use Eq. 50-1 to calculate the radii of the fragments.) (*b*) Compare this energy with the energy released in a typical fission process. In what form will this energy ultimately appear in the laboratory?

2. A ^{236}U nucleus undergoes fission and breaks up into two middle-mass fragments, ^{140}Xe and ^{96}Sr. (*a*) By what percentage does the surface area of the ^{236}U nucleus change during this process? (*b*) By what percentage does its volume change? (*c*) By what percentage does its electrostatic potential energy change? The potential energy of a uniformly charged sphere of radius r and charge Q is given by

$$U = \frac{3}{5}\left(\frac{Q^2}{4\pi\epsilon_0 r}\right).$$

3. In an atomic bomb, energy release is due to the uncontrolled fission of plutonium ^{239}Pu (or ^{235}U). The magnitude of the released energy is specified in terms of the mass of TNT required to produce the same energy release (bomb "rating"). One megaton (10^6 tons) of TNT produces 2.6×10^{28} MeV of energy. (*a*) Calculate the rating, in tons of TNT, of an atomic bomb containing 95 kg of ^{239}Pu, of which 2.5 kg actually undergoes fission. For plutonium, the average Q is 180 MeV. (*b*) Why is the other 92.5 kg of ^{239}Pu needed if it does not fission?

4. (*a*) A neutron with initial kinetic energy K makes a head-on elastic collision with a resting atom of mass m. Show that the fractional energy loss of the neutron is given by

$$\frac{\Delta K}{K} = \frac{4m_n m}{(m + m_n)^2},$$

in which m_n is the neutron mass. (*b*) Find $\Delta K/K$ if the resting atom is hydrogen, deuterium, carbon, or lead. (*c*) If $K =$ 1.00 MeV initially, how many such collisions would it take to reduce the neutron energy to thermal values (0.025 eV) if the material is deuterium, a commonly used moderator? (*Note:* In actual moderators, most collisions are not "head-on.")

5. Calculate the Coulomb barrier height for two ^7Li nuclei fired at each other with the same initial kinetic energy K. See Sample Problem 51-5. (*Hint:* Use Eq. 50-1 to calculate the radii of the nuclei.)

6. In certain stars the *carbon cycle* is more likely than the proton–proton cycle to be effective in generating energy. This cycle is

$$\begin{aligned}
{}^{12}C + {}^{1}H &\rightarrow {}^{13}N + \gamma, & Q_1 &= 1.95\ \text{MeV},\\
{}^{13}N &\rightarrow {}^{13}C + e^+ + \nu, & Q_2 &= 1.19\ \text{MeV},\\
{}^{13}C + {}^{1}H &\rightarrow {}^{14}N + \gamma, & Q_3 &= 7.55\ \text{MeV},\\
{}^{14}N + {}^{1}H &\rightarrow {}^{15}O + \gamma, & Q_4 &= 7.30\ \text{MeV},\\
{}^{15}O &\rightarrow {}^{15}N + e^+ + \nu, & Q_5 &= 1.73\ \text{MeV},\\
{}^{15}N + {}^{1}H &\rightarrow {}^{12}C + {}^{4}He, & Q_6 &= 4.97\ \text{MeV}.
\end{aligned}$$

(*a*) Show that this cycle of reactions is exactly equivalent in its overall effects to the proton–proton cycle of Fig. 51-10. (*b*) Verify that both cycles, as expected, have the same Q.

7. The gravitational potential energy of a uniform spherical object of mass M and radius R is

$$U = -3GM^2/5R,$$

in which G is the gravitational constant. (*a*) Demonstrate the consistency of this expression with that of Problem 4 in Chapter 50. (*b*) Use this expression to find the maximum energy that could be released by a spherical object, initially of infinite radius, in shrinking to the present size of the Sun. (*c*) Assume that during this shrinking, the Sun radiated energy at its present rate and calculate the age of the Sun based on the hypothesis that the Sun derives its energy from gravitational contraction.

8. (*a*) Calculate the rate at which the Sun is generating neutrinos. Assume that energy production is entirely by the proton–proton cycle. (*b*) At what rate do solar neutrinos impinge on the Earth?

9. Suppose we had a quantity of N deuterons (^2H nuclei). (*a*) Which of the following procedures for fusing these N nuclei releases more energy, and how much more? (A) $N/2$ fusion reactions of the type $^2H + {}^2H \rightarrow {}^3H + {}^1H$, or (B) $N/3$ fusion reactions of the type $^2H + {}^3H \rightarrow {}^4He + n$, using $N/3$ nuclei of ^3H that are first made in $N/3$ reactions of type A. (*b*) List the ultimate product nuclei resulting from the two procedures and the quantity of each.

10. The uncompressed radius of the fuel pellet of Sample Problem 51-8 is 20 μm. Suppose that the compressed fuel pellet "burns" with an efficiency of 10%. That is, only 10% of the deuterons and 10% of the tritons participate in the fusion reaction of Eq. 51-8. (*a*) How much energy is released in each such microexplosion of a pellet? (*b*) To how much TNT is each such pellet equivalent? The heat of combustion of TNT is 4.6 MJ/kg. (*c*) If a fusion reactor is constructed on the basis of 100 microexplosions per second, what power would be generated? (Note that part of this power must be used to operate the lasers.)

CHAPTER 52

PARTICLE PHYSICS AND COSMOLOGY

Particle physics deals with the very small. It asks questions about the structure of matter at its most fundamental or elementary level. From the debris of high-energy collisions, particle physicists have been able to create an enormous variety of new particles, study their properties, classify them into groups, and thus understand their internal structure. Cosmology deals with the very large—the origin and evolution of the universe. By studying the radiation that reaches us from throughout the universe, we can draw conclusions about how it was formed and what its future may be. It may seem surprising that we have grouped these two subjects together in one chapter. It turns out that they are closely related: what particle physicists learn about the properties of elementary particles tells us about the structure of the universe just after its birth, and conclusions by cosmologists can set limits on the varieties of particles that can exist and the interactions between them.

52-1 PARTICLE INTERACTIONS

There are tens of thousands of chemical compounds of varying degrees of complexity. Understanding this huge number of systems would be a hopeless task if it were not for the underlying simplicity of the 116 fundamental units (elements) of which these compounds are made and the relatively small number of types of bonds through which they can interact. To understand chemistry, we need not study the properties of tens of thousands of compounds, but only those of about 100 elements, along with a few basic types of bonds between them.

In fact, the task is even simpler. The 116 known elements can be classified into groups with similar properties: inert gases, halogens, alkali metals, transition metals, rare earths, and so forth. If we understand the properties of one member of a group, we can infer the properties of the other members of that group.

The subatomic world can be understood in a similar way. We know that the 116 different kinds of atoms are not fundamental units, but rather they are in turn composed of three different particles: protons, neutrons, and electrons. When we look still further, by smashing particles together at high energy and studying the debris of the collisions (see Fig. 52-1), we find what appears at first glance to be a complexity approaching that of chemistry: hundreds of different particles are produced. Yet when we look carefully we find that we can classify those particles into a few groups whose members have similar properties. Eventually we find that this classification leads to clues about the underlying substructure that is based again on a small number of truly fundamental particles and a small number of possible interactions among them.

The Four Basic Forces

All of the known forces in the universe can be grouped into four basic types. In order of increasing strength, these are: gravitation, the weak force, electromagnetism, and the strong force. These forces have important roles not only in the interactions between particles, but also in the decay of one particle into other particles.

1. *The gravitational force.* Gravity is of course exceedingly important in our daily lives, but on the scale

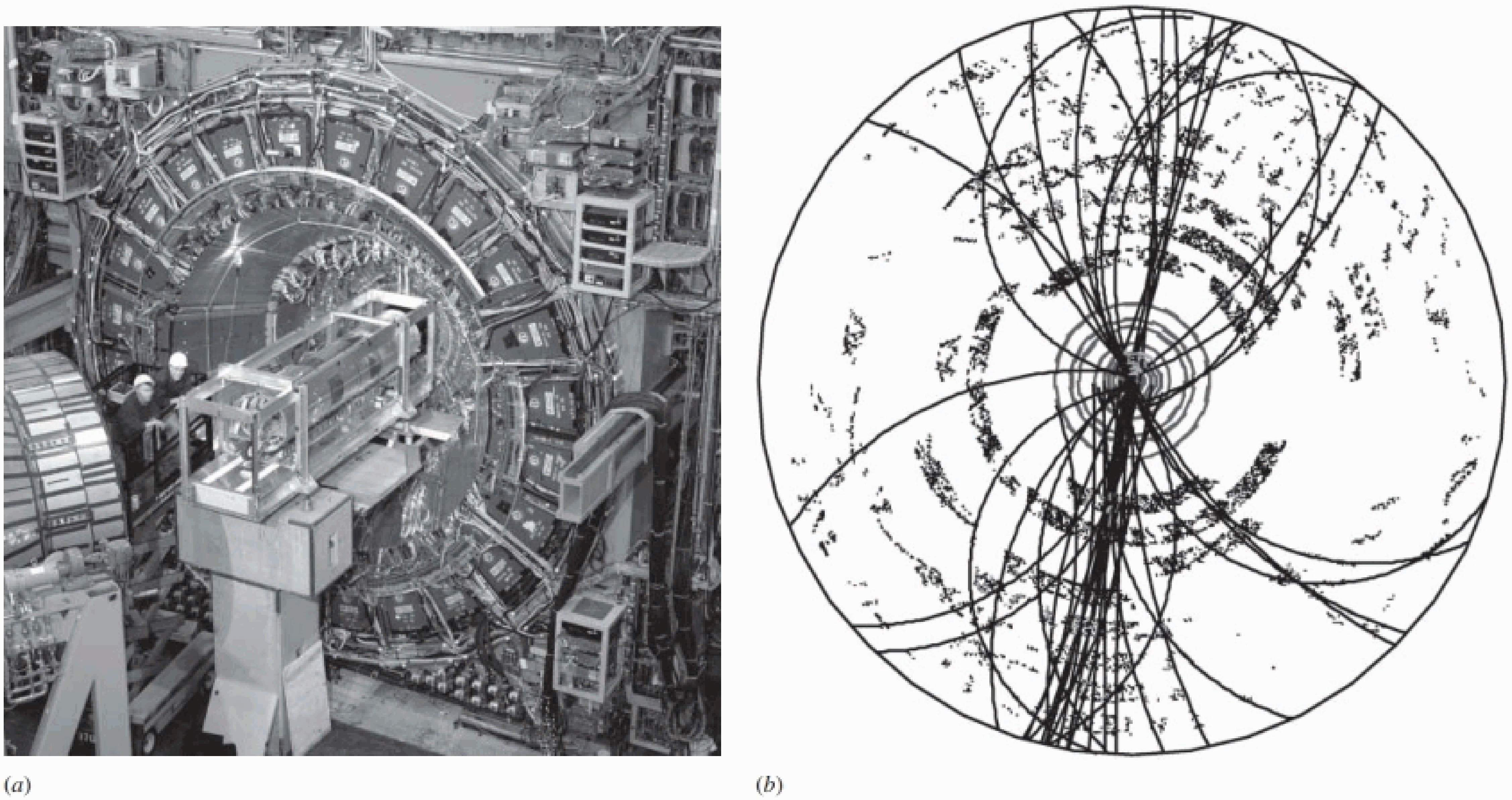


FIGURE 52-1. (*a*) The CDF detector system at Fermilab. Collisions between protons and antiprotons occur inside the detector and produce a multitude of particles, which are recorded and tracked by the detector. (*b*) An example of the trajectories of particles leaving the collision region as reconstructed by the detector. The curved paths are due to the presence of a magnetic field, which allows the momentum of the particles to be determined. For more information, see http://www-cdf.fnal.gov.

of fundamental interactions between particles in the subatomic realm, it is of no importance at all. To give a relative figure, the gravitational force between two protons just touching at their surfaces is about 10^{-38} of the strong force between them. The principal difference between gravitation and the other forces is that, on the practical scale, gravity is cumulative and infinite in range. For example, your weight is the cumulative effect of the gravitational force exerted by each atom of the Earth on each atom of your body.

2. *The weak force.* The weak force is responsible for nuclear beta decay (see Section 50-5) and other similar decay processes involving fundamental particles. It does not play a major role in the binding of nuclei. The weak force between two neighboring protons is about 10^{-7} of the strong force between them, and the range of the weak force is smaller than 1 fm. That is, at separations greater than about 1 fm, the weak force between particles is negligible. Nevertheless, the weak force is important in understanding the behavior of fundamental particles, and it is critical in understanding the evolution of the universe.

3. *The electromagnetic force.* Electromagnetism is important in the structure and the interactions of the fundamental particles. For example, some particles interact or decay primarily through this mechanism. Electromagnetic forces are of infinite range, but shielding generally diminishes their effect for ordinary objects. The properties of atoms and molecules are determined by electromagnetic forces, and many common macroscopic forces (such as friction, air resistance, drag, and tension) are ultimately due to the electromagnetic force. The electromagnetic force between neighboring protons is about 10^{-2} of the strong force, but within the nucleus the electromagnetic forces can act cumulatively because there is no shielding. As a result, the electromagnetic force can compete with the strong force in determining the stability and the structure of nuclei.

4. *The strong force.* The strong force, which is responsible for the binding of nuclei, is the dominant one in the reactions and decays of most of the fundamental particles. However, as we shall see, some particles (such as the electron) do not feel this force at all. It has a relatively short range, on the order of 1 fm.

The relative strength of a force determines the time scale over which it acts. If we bring two particles close enough together for any of these forces to act, then a longer time is required for the weak force to cause a decay or reaction than for the strong force. As we shall see, the mean lifetime of a decay process is often a signal of the type of interaction responsible for the process, with strong forces being at the shortest end of the time scale (often down to 10^{-23} s). Table 52-1 summarizes the four forces and some of their properties. The characteristic time for each force gives a typical range of time intervals observed for systems

TABLE 52-1 The Four Basic Forces

Type	Range	Relative Strength	Characteristic Time
Strong	1 fm	1	10^{-23} s
Electromagnetic	∞	10^{-2}	10^{-14}–10^{-20} s
Weak	$\ll 1$ fm	10^{-7}	10^{-8}–10^{-13} s
Gravitational	∞	10^{-38}	Years

in which each force acts. Usually this is the typical lifetime of a particle that decays through that force.

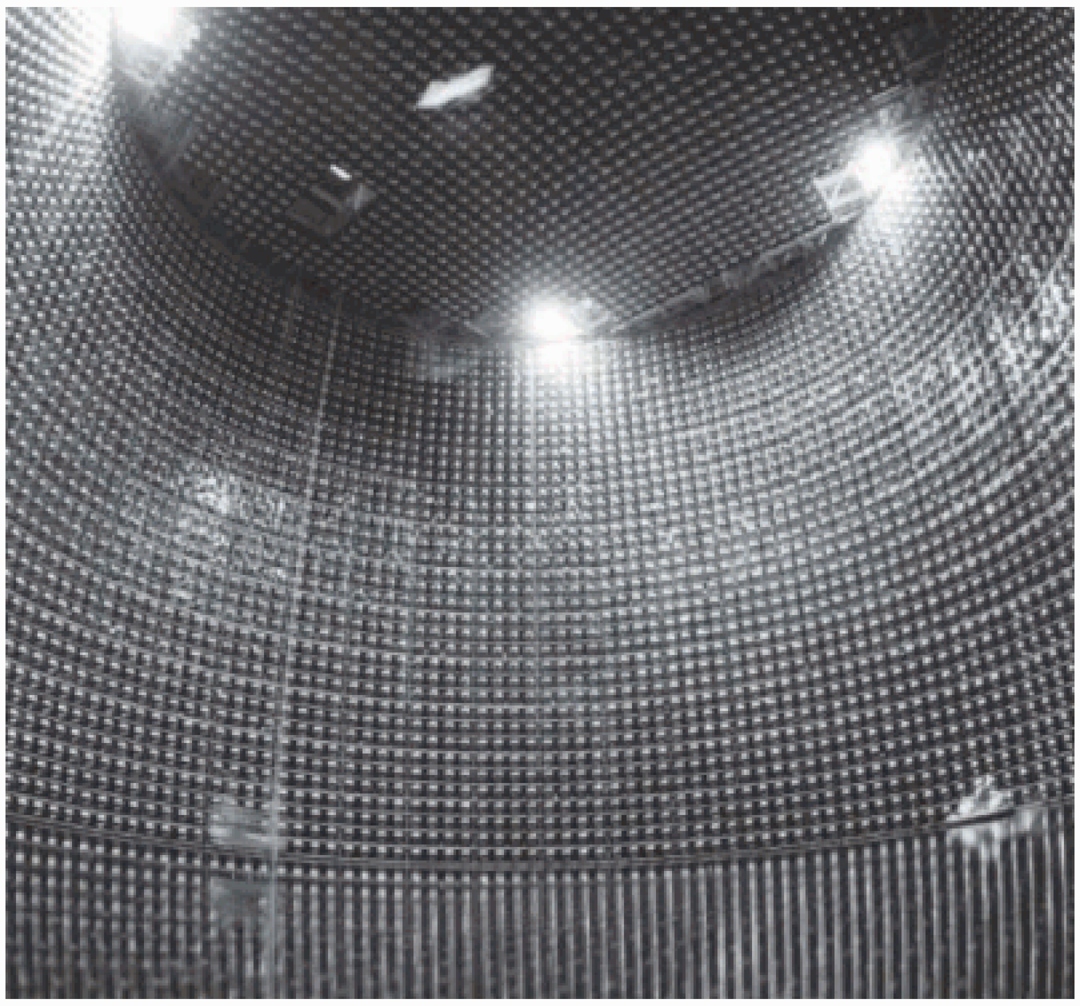


FIGURE 52-2. The Super-Kamiokande detector, located in Japan. The detector contains 32,000 tons of water. Decays of protons in the water molecules produce flashes of light, which are viewed by 11,200 photomultiplier tubes that can be seen lining the walls of the chamber. At right is a raft used by the researchers to service the tubes. The detector is also used to observe neutrinos emitted by the sun. See http://www-sk.icrr.u-tokyo.ac.jp.

Unification of Forces

One of the landmark achievements in the history of physics was the 19th century theory of electromagnetism, based on the theory of Maxwell and experiments by Faraday and Oersted showing that magnetic effects could produce electric fields and electrical effects could produce magnetic fields. The previously separate sciences of electricity and magnetism became linked under the common designation of electromagnetism. This linking was later shown to be a fundamental part of the special theory of relativity, according to which electric fields and magnetic fields can be transformed into one another due entirely to the relative motion of the observer.

In the 20th century, the attempt was made to carry this linking further to include other forces. First it was shown that electromagnetism and the weak force can be understood as two different aspects of the same force, called the *electroweak* force. If we study particle interactions at a high enough energy, these two forces behave similarly. It is convenient for us to regard them as separate forces for many of the effects we shall discuss, just as we often find it convenient to speak separately of electric and magnetic forces when we discuss electromagnetic phenomena. The theory of the electroweak force, which was proposed independently in 1967 by Stephen Weinberg and Abdus Salam (and for which they, along with Sheldon Glashow, another originator of the theory, received the 1979 Nobel Prize in physics), suggested that, just as the photon is the carrier of the electromagnetic force, there should be heavy particles that carry the weak force, and these new particles should, on an energy scale of 100 GeV (about 100 times the rest energy of the proton), behave similarly to a high-energy photon. In 1983, a research team at the European Center for Nuclear Physics (CERN), led by Carlo Rubbia and using experimental techniques developed by Simon van der Meer, discovered the predicted particles, now known as W^+, W^-, and Z^0, for which Rubbia and van der Meer were awarded the 1984 Nobel Prize in physics. The discovery of these particles provided the evidence for the unification of the electromagnetic and weak interactions into the electroweak interaction.

Next the attempt was made to combine the strong and electroweak forces at a new higher level of unification. Theories that do so are called grand unified theories (GUTs), and at the present time there are many candidates for GUTs but none has as yet emerged as the correct one. Because the energy at which the forces merge is immense, perhaps 10^{15} GeV (10^{11} times the energy of the largest particle accelerator yet built or even contemplated), we cannot do experiments to test the GUTs directly. We must therefore rely on tests at obtainable energies, where the effects are exceedingly small. One prediction of these theories is that the proton should not be a stable particle but should decay on a time scale greater than 10^{31} years. (Compare this number with the age of the universe, about 10^{10} years.) Searches for proton decay have been made by looking for a characteristic light signal that would accompany the decay of one of the protons in a large volume of water (Fig. 52-2). So far such experiments have not observed proton decay, but they have placed a lower limit of 10^{33} years on the decay lifetime. These results exclude certain of the GUTs, and experiments continue to try to verify the theories.

The final step in the unification would be to include gravity in the scheme to create a *theory of everything* (TOE). There is not yet a quantum theory of gravity, so it is difficult to anticipate the form that these theories might take, but they nevertheless provide challenges for theoretical speculation.

SAMPLE PROBLEM 52-1. Suppose the half-life of the proton were 10^{31} y, as predicted by certain GUTs. (*a*) On the average, how long must we observe a liter of water before we would see one of its protons decay? (*b*) What volume of water would be required to have a proton decay rate of one per day?

Solution (*a*) A liter of water (approximately 1000 g) contains a number of molecules given by

$$\frac{(1000\text{ g})(6.02 \times 10^{23}\text{ molecules/mole})}{18\text{ g/mole}} = 3.3 \times 10^{25}\text{ molecules.}$$

Each molecule contains 10 protons (2 from the hydrogens and 8 from the oxygen), so that the number of protons in a liter of water is $N = 3.3 \times 10^{26}$. The decay rate R is given by Eq. 50-5 as

$$R = \lambda N = \frac{\ln 2}{t_{1/2}} N = \frac{0.693}{10^{31}\text{ y}} 3.3 \times 10^{26}$$

$$= 2.3 \times 10^{-5}\text{ y}^{-1} = \frac{1}{43{,}000\text{ y}}.$$

That is, on the average we must wait for 43,000 years before a proton decay occurs in a liter of water.

(*b*) If $R = 1\text{ d}^{-1}$, we obtain

$$N = \frac{R}{\lambda} = \frac{1\text{ d}^{-1}}{0.693/(10^{31}\text{ y})} = 5.3 \times 10^{33}\text{ protons}$$

or 5.3×10^{32} molecules of water. This works out to be 1.6×10^{7} L, equivalent to a cube of water measuring 25 m on a side!

52-2 FAMILIES OF PARTICLES

We can learn a lot about things by classifying them. This is a technique used commonly by many scientists; for example, by grouping plants or animals into categories based on certain obvious features of their structure, biologists can form a basis for studying their behavior. From the scientific standpoint it may be more enlightening to compare one spider with another spider than with a fly or a moth. Part of the training of a scientist is concerned with learning how to make and to use these classifications.

The earliest classification scheme for particles was based on their masses. The lightest particles, including the electron ($m_e c^2 = 0.511$ MeV), were called *leptons* (from the Greek word for "small"). The heaviest particles, including the proton ($m_p c^2 = 938$ MeV), were called *baryons* (from the Greek word for "heavy"). In between were particles, including the pion ($m_\pi c^2 = 140$ MeV), called *mesons* (from the Greek word for "middle"). Today these classifications based on mass are no longer valid; for example, one lepton and many mesons are more massive than the proton. However, we retain these three names as descriptive of particles with similar properties, even though the classification based only on mass is no longer valid. Table 52-2 summarizes these three families of particles and some of their properties.

Leptons

The leptons are fundamental particles that interact only through the weak and electromagnetic interactions; even though the strong force can exceed the weak or electromagnetic force in strength by many orders of magnitude, the leptons do not feel this force at all. The leptons are true fundamental particles; they have no internal structure and are not composed of other, still smaller particles. We can consider the leptons to be point particles with no finite dimensions. All known leptons have a spin of $\frac{1}{2}$.

Table 52-3 shows the six leptons, which appear as three pairs of particles. Each pair includes a charged particle (e^-, μ^-, τ^-) and an uncharged neutrino (ν_e, ν_μ, ν_τ). We discussed the electron neutrino previously in connection with beta decay (Section 50-5). Both the charged leptons and the neutrinos have antiparticles.

Electron neutrinos and antineutrinos are produced in the beta decay (Section 50-5) of radioactive elements. They are also produced in great quantity in solar fusion processes; large underground detectors, such as that shown in Fig. 52-2, have been constructed to observe solar neutrinos and measure their properties. Because neutrinos interact only very weakly with matter, the solar neutrinos come to us directly from the core of the Sun, where the fusion reactions take place. (The photons from the Sun, on the other hand, come to us from its surface and therefore carry no direct information about fusion processes in the core.) Experiments have been underway for the past 40 years to count these solar neutrinos; the results of these experiments indicate that only about one-third to one-half of the expected number are reaching the Earth. Neutrinos are also produced in intense bursts from supernovas; the first such observation of supernova neutrinos was made in 1987 (Fig. 52-3).

Some theories of the properties of the neutrinos require that the neutrinos are massless, like the photon. Other theories allow a small but definitely nonzero mass. Measuring a small neutrino mass, especially for the electron neutrino, is

TABLE 52-2 Three Families of Particles

Family	*Structure*	*Interactions*	*Spin*	*Examples*
Leptons	Fundamental	Weak, electromagnetic	Half integral	e,ν
Mesons	Composite	Weak, electromagnetic, strong	Integral	π,K
Baryons	Composite	Weak, electromagnetic, strong	Half integral	p,n

TABLE 52-3 The Lepton Family

Particle	*Antiparticle*	*Particle Charge (e)*	*Spin ($h/2\pi$)*	*Rest Energy (MeV)*	*Mean Life (s)*	*Typical Decay Products*
e^-	e^+	-1	$\frac{1}{2}$	0.511	∞	—
ν_e	$\bar{\nu}_e$	0	$\frac{1}{2}$	<0.000015	∞	—
μ^-	μ^+	-1	$\frac{1}{2}$	105.7	2.2×10^{-6}	$e^- + \bar{\nu}_e + \nu_\mu$
ν_μ	$\bar{\nu}_\mu$	0	$\frac{1}{2}$	<0.19	∞	—
τ^-	τ^+	-1	$\frac{1}{2}$	1777	2.9×10^{-13}	$\mu^- + \bar{\nu}_\mu + \nu_\tau$
ν_τ	$\bar{\nu}_\tau$	0	$\frac{1}{2}$	<18	∞	—

a challenging experimental problem. So far the best experimental upper limit on the rest energy of the electron neutrino, about 15 eV, comes from the beta decay of ^{3}H. By comparing the arrival time of the neutrino burst from Supernova 1987A (Fig. 52-3) with the arrival time of the light signal from the supernova, a similar upper limit of about 20 eV was estimated.

If neutrinos do have mass, then they are permitted to transform from one type to another, such as electron neutrinos to muon neutrinos. This transformation, called *neutrino oscillation,* has not yet been seen directly, but it has been suspected as an explanation of the reduction in the number of electron neutrinos reaching us from the Sun. In 1999, the detector shown in Fig. 52-2 produced the first evidence for neutrino oscillations—the number of muon neutrinos (produced in the atmosphere by collisions of energetic cosmic rays with air molecules) coming down from above differed from the number coming up from below (that is, traveling through the Earth). This discrepancy could not be explained simply by absorption during passage through the Earth, because the neutrinos have a mean free path in matter that is measured in light-years, so they have almost no chance of being absorbed in passing through the Earth. Currently there are many experiments underway to measure neutrino masses, partly to achieve a better understanding of the properties of the neutrinos, but also because of the implications of a nonzero neutrino mass for cosmology, as we discuss later in this chapter.

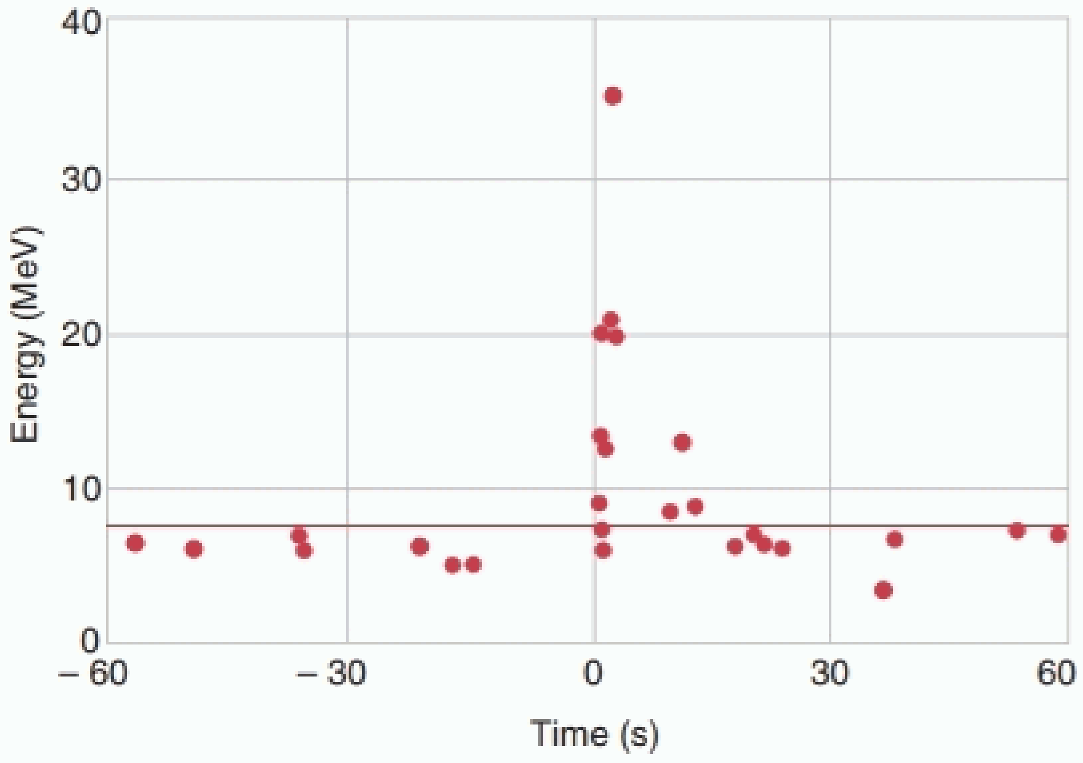


FIGURE 52-3. Evidence for a burst of neutrinos from the supernova SN 1987A.

The electron is a stable particle, but the muon and tau decay to other leptons, according to

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \qquad (\text{mean life} = 2.2 \times 10^{-6}\ \text{s}),$$

$$\tau^- \rightarrow \mu^- + \bar{\nu}_\mu + \nu_\tau \qquad (\text{mean life} = 2.9 \times 10^{-13}\ \text{s}).$$

These decays are caused by the weak interaction, as we can conclude from the presence of neutrinos (which *always* indicates a weak interaction process) among the decay products and as we infer from comparing the decay lifetimes to the characteristic times listed in Table 52-1. The form of these decays can be understood based on a conservation law for leptons discussed in Section 52-3.

Mesons

Mesons are strongly interacting particles having integral spin. A partial list of some mesons is given in Table 52-4. Generally, mesons are produced in reactions by the strong interaction; they decay, usually to other mesons or leptons, through the strong, electromagnetic, or weak interactions. For example, pions can be produced in reactions of nucleons, such as

$$p + n \rightarrow p + p + \pi^- \quad \text{or} \quad p + n \rightarrow p + n + \pi^0,$$

and the pions can decay according to

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \qquad (\text{mean life} = 2.6 \times 10^{-8}\ \text{s}),$$

$$\pi^0 \rightarrow \gamma + \gamma \qquad (\text{mean life} = 8.4 \times 10^{-17}\ \text{s}),$$

where the first decay occurs due to the weak interaction (indicated by the neutrinos and suggested by the mean life) and the second due to the electromagnetic interaction (indicated by the photons and suggested by the mean life).*

---

*Although neutrinos always indicate a weak-interaction decay, not all weak-interaction decays produce neutrinos. The same is true for photons in electromagnetic decays.

TABLE 52-4 Some Selected Mesons

Particle	*Antiparticle*	*Charge*[a] (*e*)	*Spin* ($h/2\pi$)	*Strangeness*[a]	*Rest Energy* (*MeV*)	*Mean Life* (*s*)	*Typical Decay Products*
π^+	π^-	+1	0	0	140	2.6×10^{-8}	$\mu^+ + \nu_\mu$
π^0	π^0	0	0	0	135	8.4×10^{-17}	$\gamma + \gamma$
K^+	K^-	+1	0	+1	494	1.2×10^{-8}	$\mu^+ + \nu_\mu$
K^0	$\overline{K^0}$	0	0	+1	498	0.9×10^{-10}	$\pi^+ + \pi^-$
η	η	0	0	0	547	5.5×10^{-19}	$\gamma + \gamma$
ρ^+	ρ^-	+1	1	0	770	4.4×10^{-24}	$\pi^+ + \pi^0$
η'	η'	0	0	0	958	2.2×10^{-21}	$\eta + \pi^+ + \pi^-$
D^+	D^-	+1	0	0	1869	1.1×10^{-12}	$K^- + \pi^+ + \pi^+$
ψ	ψ	0	1	0	3097	7.6×10^{-21}	$e^+ + e^-$
B^+	B^-	+1	0	0	5279	1.6×10^{-12}	$D^- + \pi^+ + \pi^+$
Υ	Υ	0	1	0	9460	1.3×10^{-20}	$e^+ + e^-$

[a] The charge and strangeness are those of the particle. Values for the antiparticle have the opposite sign. The spin, rest energy, and mean life are the same for a particle and its antiparticle.

Baryons

Baryons are strongly interacting particles having half-integral spins ($\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$). A partial listing of some baryons is given in Table 52-5. The familiar members of the baryon family are the proton and neutron. Baryons have distinct antiparticles—for example, the antiproton ($\overline{p}$) and antineutron ($\overline{n}$).

We can produce heavier baryons in reactions between nucleons, such as

$$p + p \rightarrow p + \Lambda^0 + K^+,$$

which produces the Λ^0 baryon and the K^+ meson. The Λ^0 decays according to

$$\Lambda^0 \rightarrow p + \pi^- \qquad (\text{mean life} = 2.6 \times 10^{-8}\ \text{s}).$$

Although there are no neutrinos produced in the decay, the mean life indicates that the decay is governed by the weak interaction. We shall learn the reason for this "slow" decay in Section 52-3.

Field Particles and Exchange Forces

There is one additional small family of particles that cannot be classified among the leptons, mesons, or baryons. These are the *field particles,* those responsible for carrying the forces with which the particles interact.

Newton's law of gravitation and Coulomb's law of electrostatics were originally based on the concept of "action at a distance." Later, in the nineteenth century, this concept

TABLE 52-5 Some Selected Baryons

Particle	*Antiparticle*	*Charge*[a] (*e*)	*Spin* ($h/2\pi$)	*Strangeness*[a]	*Rest Energy* (*MeV*)	*Mean Life* (*s*)	*Typical Decay Products*
p	$\overline{p}$	+1	$\frac{1}{2}$	0	938	∞	
n	$\overline{n}$	0	$\frac{1}{2}$	0	940	887	$p + e^- + \overline{\nu}_e$
Λ^0	$\overline{\Lambda}^0$	0	$\frac{1}{2}$	−1	1116	2.6×10^{-10}	$p + \pi^-$
Σ^+	$\overline{\Sigma}^+$	+1	$\frac{1}{2}$	−1	1189	0.8×10^{-10}	$p + \pi^0$
Σ^0	$\overline{\Sigma}^0$	0	$\frac{1}{2}$	−1	1193	7.4×10^{-20}	$\Lambda^0 + \gamma$
Σ^-	$\overline{\Sigma}^-$	−1	$\frac{1}{2}$	−1	1197	1.5×10^{-10}	$n + \pi^-$
Ξ^0	$\overline{\Xi}^0$	0	$\frac{1}{2}$	−2	1315	2.9×10^{-10}	$\Lambda^0 + \pi^0$
Ξ^-	$\overline{\Xi}^-$	−1	$\frac{1}{2}$	−2	1321	1.6×10^{-10}	$\Lambda^0 + \pi^-$
Δ^*	$\overline{\Delta}^*$	+2, +1, 0, −1	$\frac{3}{2}$	0	1232	5.9×10^{-24}	$p + \pi$
Σ^*	$\overline{\Sigma}^*$	+1, 0, −1	$\frac{3}{2}$	−1	1385	1.8×10^{-23}	$\Lambda^0 + \pi$
Ξ^*	$\overline{\Xi}^*$	−1, 0	$\frac{3}{2}$	−2	1530	7.3×10^{-23}	$\Xi + \pi$
Ω^-	$\overline{\Omega}^-$	−1	$\frac{3}{2}$	−3	1672	8.2×10^{-11}	$\Lambda^0 + K^-$
Λ_c^+	$\overline{\Lambda}_c^+$	+1	$\frac{1}{2}$	0	2285	2.1×10^{-13}	$p + K^- + \pi^+$
Λ_b^0	$\overline{\Lambda}_b^0$	0	$\frac{1}{2}$	0	5624	1.2×10^{-12}	$p + D^0 + \pi^-$

[a] The charge and strangeness are those of the particle. Values for the antiparticle have the opposite sign. The spin, rest energy, and mean life are the same for a particle and its antiparticle.

TABLE 52-6 The Field Particles

Particle	*Symbol*	*Interaction*	*Charge (e)*	*Spin ($h/2\pi$)*	*Rest Energy (GeV)*
Graviton		Gravitation	0	2	0
Weak boson	W^+, W^-	Weak	± 1	1	80.4
Weak boson	Z^0	Weak	0	1	91.2
Photon	γ	Electromagnetic	0	1	0
Gluon	g	Strong (color)	0	1	0

was replaced by the notion of a field. Two particles interact through the fields that they establish: one particle sets up a field, and the other interacts with that field rather than directly with the first particle. Quantum field theory takes this notion one step further by supposing that the fields are carried by quanta. In this view, instead of the first particle setting up the field, we say that it emits quanta of the field. The second particle then absorbs these quanta. For example, the electromagnetic interaction between two particles can be viewed in terms of the emission and absorption of photons, which are quanta of the electromagnetic field. Each type of field has its characteristic field particles. A list of the particles associated with the four basic forces can be found in Table 52-6.

A force accomplished through the exchange of particles is called an *exchange force*. For example, the force between two nucleons in a nucleus takes place through the exchange of pions. In this case the pions, along with other mesons, can act as field particles associated with the strong force between nucleons.

How is it possible for a particle, such as a proton, to emit another particle with nonzero mass and still remain a proton? This process seems to violate conservation of energy. The solution to this dilemma lies in the energy–time form of the uncertainty relationships. The uncertainty principle is a fundamental limitation on our ability to measure a system. That is, if we observe a system for a time interval Δt, there is a corresponding uncertainty ΔE in its energy, according to Eq. 46-9, given at minimum by

$$\Delta E = \frac{h}{2\pi\,\Delta t}. \qquad (52\text{-}1)$$

We cannot know the energy of a system more precisely than this ΔE unless we measure for a time longer than Δt. If we observe only for a very short time, the uncertainty in the rest energy of a proton can be at least as large as the rest energy of a pion, as the following sample problem demonstrates.

SAMPLE PROBLEM 52-2. (*a*) What is the longest interval of time for which we can observe a proton for the uncertainty in its rest energy to be equal to the pion rest energy? (*b*) What is the greatest distance the pion can travel in that time?

Solution (*a*) For the proton's rest energy to be uncertain by an amount $\Delta E = m_\pi c^2$, the observation time interval can, according to Eq. 52-1, be at most

$$\Delta t = \frac{h}{2\pi\,\Delta E} = \frac{h}{2\pi m_\pi c^2}$$
$$= \frac{4.14 \times 10^{-15}\ \text{eV}\cdot\text{s}}{(2\pi)(140\ \text{MeV})} = 4.7 \times 10^{-24}\ \text{s}.$$

In a time interval shorter than 4.7×10^{-24} s, a proton can emit and absorb a pion without our observing a violation of conservation of energy.

(*b*) If the pion travels at nearly the speed of light, the maximum distance d it can travel in this time interval is

$$d = c\,\Delta t = (3.00 \times 10^8\ \text{m/s})(4.7 \times 10^{-24}\ \text{s})$$
$$= 1.4 \times 10^{-15}\ \text{m} = 1.4\ \text{fm}.$$

This distance defines the *range* of the nuclear force. Two nucleons closer than about 1.4 fm can interact through the exchange of pions. If the nucleons are separated by a greater distance, pion exchange cannot operate, and there is no nuclear force.

52-3 CONSERVATION LAWS

We would have a difficult time analyzing physical processes without the laws of conservation of energy and linear and angular momentum. These conservation laws help us understand why certain outcomes occur (such as in the case of the collisions that we considered in Chapter 6). They also help us understand why certain processes (those that violate the conservation laws) are never observed. In one sense they are empirical laws, deduced from observing physical processes and carefully tested in the laboratory. In another sense they reveal to us fundamental aspects of the laws of nature.

An example of a conservation law is the conservation of electric charge. By observing the outcomes of many processes, we are led to propose this law: the net amount of electric charge must not change in any process. Equivalently, we may say that the net charge before a particular reaction or decay must equal the net charge after the reaction or decay. No violation of this law has ever been observed, even though it has been carefully tested (see Section 25-6).

Conservation of Lepton Number

In reactions and decays of fundamental particles, we often find a certain set of outcomes but fail to observe a set of related outcomes that would otherwise be expected to occur.

When this happens, we suspect that there is some unknown conservation law at work that permits the first set and forbids the second. For example, we can produce an electron neutrino when a proton captures an electron:

$$e^- + p \rightarrow n + \nu_e.$$

We always find neutrinos in this process, but we never observe antineutrinos. Furthermore, the reaction always produces electron neutrinos and never muon or tau neutrinos.

We account for the failure to observe certain processes by proposing a conservation law for *lepton number* that is similar to the conservation law for electric charge. To each lepton we assign a lepton number $+1$ and to each antilepton we assign a lepton number -1. All other particles have lepton numbers of 0. The law of conservation of lepton number then states:

> *In any process, the lepton number for electron-type leptons, muon-type leptons, and tau-type leptons must each remain constant.*

As far as we know, the law of lepton conservation is strictly valid; despite careful experimental searches for violations, none has yet been found.

In the electron capture process, we assign an electron lepton number L_e of $+1$ to the electron and to the electron neutrino, whereas $L_e = 0$ for the proton and neutron. This process then has $L_e = +1$ on both sides and upholds the law of conservation of lepton number. If an electron *anti*neutrino were produced, the right side would have $L_e = -1$, and the law would be violated. This accounts for our failure to observe this process. If another type of neutrino—for example, a muon neutrino—were produced, the process would have $L_e = +1$ on the left and $L_e = 0$ on the right. Furthermore, it would have $L_\mu = 0$ on the left and $L_\mu = +1$ on the right. The process would therefore violate conservation of both electron and muon lepton numbers, and it has never been observed.

Through the law of lepton conservation, we can account for many experimental observations. Like other conservation laws, this law proves to be of great value in analyzing decays and reactions.

SAMPLE PROBLEM 52-3. Use lepton number conservation to determine the identity of particle X in the following decay: $\mu^+ \rightarrow e^+ + \nu_e + X$.

Solution We assign electron-type lepton number L_e and muon-type lepton number L_μ to the particles as follows:

$$\begin{array}{lcccc} & \mu^+ \rightarrow & e^+ + & \nu_e + & X \\ L_e: & 0 & -1 & +1 & ? \\ L_\mu: & -1 & 0 & 0 & ? \end{array}$$

Note that electron-type leptons are assigned $L_\mu = 0$ and muon-type leptons are assigned $L_e = 0$.

Conservation of lepton number means that the net electron-type lepton number must be the same before and after the decay, and similarly for the muon-type lepton number. The electron-type lepton number is 0 before the decay and so must be 0 after the decay, which requires that particle X have $L_e = 0$. The muon-type lepton number is -1 before the decay and to make the net muon-type lepton number equal -1 after the decay, particle X must have $L_\mu = -1$.

Particle X must therefore be a lepton with $L_e = 0$ and $L_\mu = -1$. Because the other particles in the decay already satisfy conservation of electron charge, particle X must be uncharged. A glance at Table 52-3 shows that the only particle with these characteristics is the muon-type antineutrino $\bar{\nu}_\mu$.

Conservation of Baryon Number

A similar conservation law occurs in the case of baryons. To each baryon, such as the proton or neutron, we assign a baryon number B of $+1$, and we assign $B = -1$ to antibaryons such as the antiproton. The law of conservation of baryon number then states:

> *In any process, the total baryon number must remain constant.*

No violation of this law has yet been observed. (However, certain speculative theories, the GUTs discussed in Section 52-1, suggest that the proton can decay into nonbaryons, which would violate the law of conservation of baryon number. This decay has never been observed; if it were observed, the law of conservation of baryon number would need to be changed accordingly.)

Consider, for example, the reaction in which antiprotons are produced when a proton beam is incident on a target of protons:

$$\begin{array}{lcccccccc} & p & + & p & \rightarrow & p & + & p & + & p & + & \bar{p} \\ B: & +1 & & +1 & & +1 & & +1 & & +1 & & -1 \end{array}$$

In this reaction, the net baryon number is $+2$ on both the left and the right sides.

Contrary to the case of lepton number, there is only one type of baryon number. The law of conservation of baryon number is a more general version of the rule we used in analyzing nuclear processes in Chapters 50 and 51; there we kept the total of neutrons plus protons constant in all decays and reactions, which, because neutrons and protons are baryons, is equivalent to conserving the total number of baryons.

Even though there are conservation laws for two types of particles (leptons and baryons), there is no conservation law for mesons. For example, in a reaction of protons on protons, any number of mesons can be produced (as long as the incident particles have enough kinetic energy):

$$p + p \rightarrow p + n + \pi^+,$$

$$p + p \rightarrow p + p + \pi^+ + \pi^-,$$

$$p + p \rightarrow p + n + \pi^+ + \pi^0 + \pi^0.$$

Note the conservation of electric charge in these processes.

Strangeness

There are still other processes that are difficult to understand based only on the conservation laws we have discussed so far. For example, consider the group of kaons (K mesons), which in many respects are similar to the pions. Because there is no conservation law for mesons, we might expect that any number of kaons can be produced in reactions. What we instead find is that kaons are either produced in pairs, for example,

$$p + p \rightarrow p + p + K^+ + K^-,$$
$$p + p \rightarrow p + n + K^+ + \overline{K^0},$$

or if a single kaon is produced, it is always accompanied by another "strange" particle, for example, a Λ^0,

$$p + p \rightarrow p + \Lambda^0 + K^+.$$

We account for these processes (and the failure to observe others that appear to be permitted by the previously known conservation laws) by assigning to particles a new quantum number called *strangeness,* which is found to follow a new conservation law, called *conservation of strangeness.* Two kaons (K^+ and K^0) are assigned to have strangeness $S = +1$, and the other two (K^- and $\overline{K^0}$) are assigned $S = -1$. All nonstrange particles (such as p, n and e) have $S = 0$. The reactions in which two kaons are produced then have $S = 0$ on the left (only nonstrange particles) and also $S = 0$ on the right. The Λ^0 baryon is assigned $S = -1$, so the reaction in which $\Lambda^0 + K^+$ is produced also has $S = 0$ on both sides.

When we analyze the decays of the strange particles, the conservation of strangeness appears occasionally to break down. The kaons can decay into two (nonstrange) pions, for example,

$$K^+ \rightarrow \pi^+ + \pi^0.$$

Here we have $S = +1$ on the left and $S = 0$ on the right, a clear violation of the conservation of strangeness. We get a clue about how to resolve this difficulty when we measure the lifetime for this decay, which turns out to be about 10^{-8} s. The kaons and pions are strongly interacting particles, and we would expect this decay to occur with a typical strong interaction lifetime in the range of 10^{-23} s (see Table 52-1). Instead, it is slowed by 15 orders of magnitude! What could be responsible for slowing this decay?

Another clue comes from the more commonly observed decay mode of the K^+:

$$K^+ \rightarrow \mu^+ + \nu_\mu,$$

a weak-interaction process, for which the mean life of 10^{-8} s would not be unusual. It appears that *the weak interaction can change strangeness* by one unit. In either of these kaon decay modes, we have S changing by one unit. Even though it does not produce the neutrinos that usually characterize a weak-interaction process, the decay $K^+ \rightarrow \pi^+ + \pi^0$ is governed by the weak interaction. In this case, the strangeness violation is a clue that it cannot be a strong-interaction process (strangeness is conserved in all strong interactions), and it must therefore be a weak-interaction decay.

Does the electromagnetic interaction conserve strangeness? To answer this question, we look for strangeness-violating electromagnetic decays, such as $\Lambda^0 \rightarrow n + \gamma$. This decay apparently does not occur, and so we conclude that *the electromagnetic interaction conserves strangeness.*

We can summarize these results in the law of conservation of strangeness:

> *In processes governed by the strong or electromagnetic interactions, the total strangeness must remain constant. In processes governed by the weak interaction, the total strangeness either remains constant or changes by one unit.*

SAMPLE PROBLEM 52-4. The Ω^- baryon has $S = -3$. (*a*) It is desired to produce the Ω^- using a beam of K^- incident on protons. What other particles are produced in this reaction? (*b*) How might the Ω^- decay?

Solution (*a*) Reactions usually proceed only through the strong interaction, which conserves strangeness. We consider the reaction

$$K^- + p \rightarrow \Omega^- + ?$$

On the left side, we have $S = -1$, $B = +1$, and $Q = 0$. On the right side, we have $S = -3$, $B = +1$, and $Q = -1$. We must therefore add to the right side particles with $S = +2$, $B = 0$, and $Q = +1$. Scanning through the tables of mesons and baryons, we find that we can satisfy these criteria with K^+ and K^0, so the reaction is

$$K^- + p \rightarrow \Omega^- + K^+ + K^0.$$

(*b*) The Ω^- cannot decay by the strong interaction, because no $S = -3$ final states are available. It must therefore decay to particles having $S = -2$ through the weak interaction, which can change S by one unit. One of the product particles must be a baryon in order to conserve baryon number. Two possibilities are

$$\Omega^- \rightarrow \Lambda^0 + K^- \quad \text{and} \quad \Omega^- \rightarrow \Xi^0 + \pi^-.$$

52-4 THE QUARK MODEL

Decays and reactions involving mesons and baryons are subject to conservation laws involving two quantities: the electric charge Q and the strangeness S. It then makes sense to ask whether there is any connection between the electric charge and the strangeness of a particle. In a particular group of similar particles (the spin-0 mesons or the spin-$\frac{1}{2}$ baryons, for example), do we find all possible combinations of Q and S or only certain ones? Finding only a restricted set of combinations suggests that the particles are built according to a set of rules out of more fundamental units whose electric charge and strangeness have certain values.

To begin to answer the question about the possible internal structure of particles, we examine the connection between electric charge and strangeness by making a plot showing electric charge on one axis and strangeness on

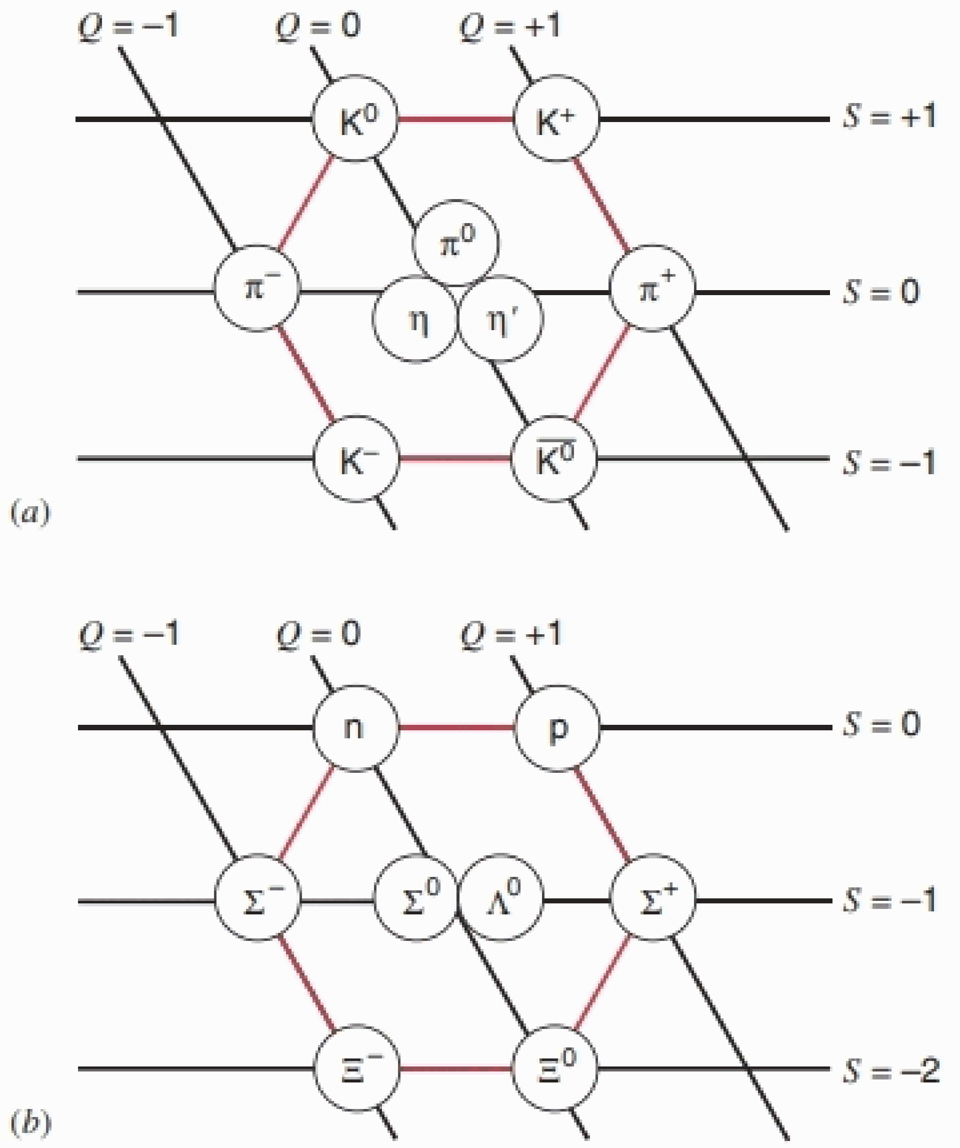


FIGURE 52-4. A chart showing (*a*) the spin-0 mesons and (*b*) the spin-$\frac{1}{2}$ baryons. Each particle is located on a grid according to its strangeness S and electric charge Q. The grid lines for electric charge have been drawn obliquely so that the patterns appear more symmetric.

another. Placing a group of nine of the spin-0 mesons (from Table 52-4) on this grid, we obtain Fig. 52-4*a*; doing the same for the spin-$\frac{1}{2}$ baryons (Table 52-5), we obtain Fig. 52-4*b*. The regularity of these patterns suggests that there may be an underlying common structure to these particles.

In 1964, it was suggested independently by Murray Gell-Mann and George Zweig that these regular patterns could be explained if it were assumed that the mesons and baryons are composed of a set of more fundamental particles called *quarks.* The original model proposed by Gell-Mann and Zweig consisted of three quarks, which are called up (u), down (d), and strange (s), along with their antiparticles (the antiquarks $\bar{\mathrm{u}}$, $\bar{\mathrm{d}}$, and $\bar{\mathrm{s}}$). Their properties are listed in Table 52-7. Later it was discovered that three quarks are not sufficient to explain the properties of all mesons and baryons; a total of six quarks is needed, as we discuss below.

The quarks have two very unusual properties that make them different from other particles: (1) They have fractional electric charges, equal to (in units of the elementary charge e) $+\frac{2}{3}$ for the u quark and $-\frac{1}{3}$ for the d and s quarks. No other particle has a fractional electric charge, and no experiment has ever confirmed the existence of free particles with fractional charge. (2) The quarks also have fractional baryon numbers; all known free particles have baryon numbers of $+1$ (for baryons), -1 (for antibaryons), or 0 (for nonbaryons such as mesons or leptons). The electric charge and the baryon number for antiquarks have the same magnitude but the opposite sign as the values for the corresponding quark.

The quarks all have a spin of $\frac{1}{2}$, which means that (according to the rules for combining spins in quantum mechanics) combinations of two quarks (or a quark and an antiquark) can have a total spin of 0 or 1, and combinations of three quarks can have a total spin of $\frac{1}{2}$ or $\frac{3}{2}$. For the lighter mesons and baryons discussed here, we assume that the combinations of quarks have no orbital angular momentum; that is, in the language of Chapter 48, they combine in *s* states ($l = 0$). There are indeed combinations of quarks that involve *p* states ($l = 1$) or *d* states ($l = 2$), but they are more massive than the particles we are considering.

According to this model, the mesons are composed of a quark and an antiquark, and the baryons are composed of three quarks. Consider the combination $\mathrm{u\bar{d}}$ of an up quark and an antidown quark, such that their two spins add to 0. The charge of the up quark (in units of e) is $+\frac{2}{3}$, and the charge of the antidown quark is $+\frac{1}{3}$ (the charge of an antiparticle is opposite to that of the particle). The combination $\mathrm{u\bar{d}}$ has $Q = +1$, $S = 0$ (because both quarks have $S = 0$), and $B = 0$ (because the quark has $B = +\frac{1}{3}$ and the antiquark has $B = -\frac{1}{3}$). This combination has the same quantum numbers as the π^+ meson. Continuing in this way, we find nine possible combinations of a quark and an antiquark, which are listed in Table 52-8. These nine combinations exactly reproduce the electric charge and strangeness combinations of the spin-0 mesons.

Baryons are composed of three quarks. Since each of the quarks has a baryon number of $+\frac{1}{3}$, the total baryon number of these composite particles will be $+1$, as we expect for baryons. Consider for example the combination uud with spin $\frac{1}{2}$. The total electric charge is $Q = +\frac{2}{3} + \frac{2}{3} - \frac{1}{3} = +1$, and the total strangeness is 0. This matches the properties of the proton. There are a total of nine possible combinations of three quarks, which are listed in Table 52-9 along with the identification of the corresponding baryon from Table 52-5. Once again, the model has successfully accounted for the properties of this group of particles. Similar success is obtained with the spin-$\frac{3}{2}$ baryons (see Exercises 14 and 15).

TABLE 52-7 **Properties of Three Quarks**

Quark	*Symbol*	*Antiquark*	*Charge*[a] *(e)*	*Spin* *(h/2π)*	*Baryon Number*[a]	*Strangeness*[a]
Up	u	$\bar{\mathrm{u}}$	$+\frac{2}{3}$	$\frac{1}{2}$	$+\frac{1}{3}$	0
Down	d	$\bar{\mathrm{d}}$	$-\frac{1}{3}$	$\frac{1}{2}$	$+\frac{1}{3}$	0
Strange	s	$\bar{\mathrm{s}}$	$-\frac{1}{3}$	$\frac{1}{2}$	$+\frac{1}{3}$	−1

[a] The values for charge, baryon number, and strangeness refer to the quarks. Values for the antiquarks have opposite signs.

TABLE 52-8 Quark–Antiquark Combinations

Combination	*Charge (e)*	*Baryon Number*	*Strangeness*	*Identity*
$u\bar{u}$	0	0	0	π^0, η, η'
$d\bar{d}$	0	0	0	
$s\bar{s}$	0	0	0	
$u\bar{d}$	+1	0	0	π^+
$u\bar{s}$	+1	0	+1	K^+
$d\bar{u}$	−1	0	0	π^-
$d\bar{s}$	0	0	+1	K^0
$s\bar{u}$	−1	0	−1	K^-
$s\bar{d}$	0	0	−1	$\overline{K^0}$

The quark model does far more than reproduce the simple geometrical patterns of Fig. 52-4. You should think of these patterns as ways of organizing particles with similar properties, just as the periodic table allows us to organize atoms with similar properties. Underlying the periodic table is atomic theory, which can be used to calculate properties of atoms beyond their geometrical arrangements. In a similar way, the quark model allows us to calculate properties of particles, including masses, magnetic dipole moments, decay modes, lifetimes, and reaction products. The agreement between the measured and calculated properties has been a spectacular success for the model. In fact, all known particles (hundreds of them!) have been accounted for based on this model, with a few additional quarks that we describe later.

The most unusual aspect of the quark model is the fractional electric charges of the quarks. All particles yet discovered have electric charges that can be expressed as integral multiples of the basic unit of charge e. No particle with a fractional electric charge has ever been seen. In fact, no one has ever seen a free quark, despite heroic experimental efforts to search for one. It is possible that our particle accelerators do not yet have enough energy to produce a free quark. It has also been suggested that free quarks may be forbidden to exist, so we may only observe quarks bound in mesons and baryons.

Even though free quarks have never been seen, individual *bound* quarks have been observed. Scattering experiments that probe deep inside the nucleon have revealed three pointlike objects that appear to have a spin of $\frac{1}{2}$ and a charge of $+\frac{2}{3}$ or $-\frac{1}{3}$. These experiments give direct proof of the existence of quarklike particles within the nucleon.

The Interaction between Quarks

What holds the quarks together inside a meson or a nucleon? This force is the most fundamental version of the strong force, brought about through the exchange of particles called *gluons.* Just as the electromagnetic interaction between charged particles can be regarded as an exchange of photons, the strong interaction between quarks is accomplished through the exchange of gluons. We therefore picture a nucleon as composed of three quarks mutually exchanging gluons.

The interaction between quarks has two unusual properties. (1) It takes a large (perhaps infinite) energy to separate two quarks to a distance greater than the size of a nucleon or a meson (about 1 fm). This may be the reason that no free quarks have yet been seen. When we try to pump energy into a nucleon to separate one of its quarks, the energy actually creates a quark–antiquark pair. The antiquark combines with one of the quarks to form a meson, which agrees with our

TABLE 52-9 Combinations of Three Quarks

Combination	*Charge (e)*	*Spin ($h/2\pi$)*	*Baryon Number*	*Strangeness*	*Spin $\frac{1}{2}$ Identity*
uuu	+2	$\frac{3}{2}$	+1	0	—
uud	+1	$\frac{3}{2}, \frac{1}{2}$	+1	0	p
udd	0	$\frac{3}{2}, \frac{1}{2}$	+1	0	n
uus	+1	$\frac{3}{2}, \frac{1}{2}$	+1	−1	Σ^+
uss	0	$\frac{3}{2}, \frac{1}{2}$	+1	−2	Ξ^0
uds	0	$\frac{3}{2}, \frac{1}{2}$	+1	−1	Λ^0, Σ^0
ddd	−1	$\frac{3}{2}$	+1	0	—
dds	−1	$\frac{3}{2}, \frac{1}{2}$	+1	−1	Σ^-
dss	−1	$\frac{3}{2}, \frac{1}{2}$	+1	−2	Ξ^-
sss	−1	$\frac{3}{2}$	+1	−3	—

observations: when we smash two nucleons together at high energies, we get our nucleons (or other baryons) back, plus some additional mesons. The more energy we put in, the more mesons we get out, but no free quark emerges. (2) Paradoxically, inside the nucleon or the meson, the quarks appear to move freely. At very short distances (less than the size of a nucleon), the force between quarks approaches zero.

This unusual behavior of quarks and gluons can be understood by comparison with electromagnetism. Two charged particles interact with one another through the exchange of photons. However, the photon itself carries no electric charge, and so the interaction between the charged particle and the exchanged photon does not result in the exchange of additional photons. A quark, on the other hand, can emit a gluon and interact with it. This interaction between the quark and the gluon can create additional gluons. When it interacts with another electron, an electron can emit a photon and still remain an electron. It does not sacrifice its "electricness" (that is, its electric charge) to emit the photon. A quark, however, gives its emitted gluon a share of its "strongness," which physicists call "color." In the interaction of quarks, color plays the same role as electric charge in the interaction of charged particles. A photon carries no electric charge, but a gluon carries color, and in doing so it changes the residual color left behind in the quark that emitted the gluon. In effect, the quark is spreading its color over a sphere the size of a nucleon (the range of the gluons), and as a result the interaction between quarks is considerably weakened at these distances.

Particle physicists have chosen amusing and whimsical names to describe the fundamental particles and their properties. Names such as quark, strangeness, gluon, or color have meaning only as labels. Gluons do provide the "glue" that binds quarks together, but it has no similarity to any other "glue" in our experience. The "color" carried by quarks and gluons has nothing to do with our ordinary use of color. It is simply easier to remember and discuss these properties if we give them familiar names.

More Quarks

In simultaneous 1974 experiments at the Brookhaven National Laboratory in New York and the Stanford Linear Accelerator Center in California, investigators discovered an unusual meson with a rest energy about three times that of the proton. This new meson, called ψ (psi), was expected to decay into lighter mesons in a strong interaction time of perhaps 10^{-23} s. Instead, it was observed to decay in a time of about 10^{-20} s, which is more characteristic of the electromagnetic interaction (see Table 52-1). Moreover, its decay products were not mesons but an electron and a positron, another signal of an electromagnetic process.

Why is the rapid, strong-interaction decay path blocked for this particle, slowing its decay by three orders of magnitude? We discussed a similar effect in the case of strangeness, a new quantum number that was introduced partly to explain certain slow decays. We accounted for those decays through a violation of the conservation of strangeness.

In a similar fashion, we assume that the decay of ψ is slowed by the violation of another conservation law, called *charm.* According to this interpretation, the ψ meson is composed of a new quark c (for charm) and its antiquark $\overline{\mathrm{c}}$. The c quark has an electric charge of $+\frac{2}{3}$. Just as the strange quark is assigned a strangeness quantum number of $S = -1$, the charmed quark is assigned a charm of $C = +1$. The decay of the ψ meson is slowed, because the c quark must decay into other quarks (u, d, or s), all of which have $C = 0$. The decay thus involves a violation of the conservation of charm and therefore cannot occur through the strong interaction, which conserves charm. The c quark can also combine with other quarks to make a meson or a baryon; for example, the D^+ meson listed in Table 52-4 has quark content $\mathrm{c}\overline{\mathrm{d}}$, and the Λ_c^+ baryon listed in Table 52-5 has quark content udc.

In 1977, a similar discovery was made at the Fermi National Accelerator Laboratory near Chicago. Again, a heavy meson (in this case, ten times the proton rest energy) was discovered, which was expected to decay to other mesons in a time characteristic of the strong interaction, but instead it decayed into $\mathrm{e}^- + \mathrm{e}^+$ in about 10^{-20} s. In this case, the decay was again slowed by the violation of yet another conservation rule, involving yet another new quark, called b (for bottom) and having an electric charge of $-\frac{1}{3}$. This new meson, called Υ (upsilon), is assumed to be composed of the combination $\mathrm{b}\overline{\mathrm{b}}$. If we assign to the b quark a new quantum number that represents bottomness, then the decay is slowed because the b quark must change into lighter quarks that lack this property; this violation of the conservation of bottomness is responsible for slowing the decay. The b quark can also combine with other quarks to make various mesons and baryons; for example, the B^+ meson listed in Table 52-4 has quark content $\mathrm{u}\overline{\mathrm{b}}$, and the Λ_b^0 baryon listed in Table 52-5 is composed of udb quarks.

In 1994, another quark was discovered from proton–proton collisions at Fermilab.* These high-energy collisions created a new quark and its antiquark, which rapidly decayed into a shower of secondary particles; from the momentum of these secondary particles, the existence of the original decaying particles was deduced. This new quark, with a charge of $+\frac{2}{3}$, is called the top quark t, and like the charmed and bottom quarks it has an associated property called *topness.*

A New Symmetry

Ordinary matter is composed of protons and neutrons, which are in turn made up only of u and d quarks. Ordinary matter is also composed of electrons, and in the conversion of protons to neutrons or neutrons to protons in the beta decay of ordinary matter, we find electron-type neutrinos emitted along with the positron or electron.

---

*For details of this discovery, see http://www.fnal.gov/pub/top95/top95_background.html.

We can therefore construct our entire world and all the phenomena we commonly observe out of two pairs of fundamental particles: u and d quarks, and e^- and ν_e leptons. Within each pair, the charges differ by one unit ($+\frac{2}{3}$ and $-\frac{1}{3}$; -1 and 0).

If we do experiments at a somewhat higher energy, we find new types of particles: a new pair of leptons (μ^- and its neutrino ν_μ) and a new pair of quarks (c and s). Once again, within each pair the electric charges differ by one unit. At still higher energies, we find a new pair of leptons (τ and ν_τ) and a new pair of quarks (t and b).

It therefore seems that the truly fundamental particles, the quarks and leptons, appear in pairs, and that a pair of quarks and a pair of leptons can be combined into a "generation," as follows:

$$\text{1st generation:} \quad \begin{pmatrix} \mathrm{e} \\ \nu_e \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathrm{u} \\ \mathrm{d} \end{pmatrix}$$

$$\text{2nd generation:} \quad \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathrm{c} \\ \mathrm{s} \end{pmatrix}$$

$$\text{3rd generation:} \quad \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathrm{t} \\ \mathrm{b} \end{pmatrix}$$

Properties of these six quarks and leptons are summarized in Appendix F.

The model of the structure of the elementary particles that we have described so far has come to be known as the *Standard Model.* In this model, there are three generations of leptons and quarks, which are considered to be elementary particles. All other material particles are composites. The particles interact through the electroweak and the strong forces (gravitation having a negligible effect on the interactions of individual particles). The electroweak force is due to the exchange of photons or the so-called weak bosons W^+, W^-, and Z^0. The six quarks each come in three varieties or colors; there are eight colored gluons that carry the color charge and are responsible for the interaction between quarks. All baryons and mesons are colorless; the three quarks in a baryon must each have a different color, and the quark and antiquark in a meson must carry one color and its anticolor.

Although the Standard Model has had remarkable success in accounting for the properties and interactions of particles, you may now be wondering if we have simply replaced one level of complexity with another. Is it possible that, instead of hundreds of mesons and baryons, our list of particles might show hundreds of leptons and quarks as we do experiments at ever greater energies? There is very strong evidence that the number of generations is exactly three, and that the list of leptons and quarks we have described is complete. One limit on the number of generations of leptons comes from the decay of certain particles. For example, the Z^0 can decay into a lepton and its antilepton, such as $e^+ + e^-$ or $\mu^+ + \mu^-$. The lifetime for the decay is determined by how many of these pairs into which it can decay, and from the observed lifetime it is concluded that the number of lepton generations can be no greater than three.

A very different approach to answering this question comes from examining the properties of the universe a short time after its birth. At that time, the universe was a hot gas consisting of leptons and quarks. The evolution and subsequent properties of the universe were determined by the number of generations that existed at that time. From observations of the present properties of the universe, we can again conclude that the number of generations is again three. It is therefore very likely that our list of leptons and quarks is complete.

SAMPLE PROBLEM 52-5. Analyze these processes in terms of their quark content:

$$(a) \quad \mathrm{p} \rightarrow \mathrm{n} + e^+ + \nu_e,$$

$$(b) \quad \Omega^- \rightarrow \Lambda^0 + \mathrm{K}^-,$$

$$(c) \quad \mathrm{K}^- + \mathrm{p} \rightarrow \Omega^- + \mathrm{K}^+ + \mathrm{K}^0.$$

Solution (*a*) Using Table 52-9 to find the quark content of the particles, we can rewrite the decay as

$$\mathrm{uud} \rightarrow \mathrm{udd} + e^+ + \nu_e.$$

Canceling the common pair of ud quarks from each side, we find

$$\mathrm{u} \rightarrow \mathrm{d} + e^+ + \nu_e.$$

The u quark changes to a d quark by beta decay. This equation represents a fundamental interaction between quarks which we can also write as

$$\mathrm{u} \rightarrow \mathrm{d} + W^+ \quad \text{and} \quad W^+ \rightarrow e^+ + \nu_e.$$

(*b*) The quark content of the Ω^- is sss (because it has strangeness -3), so the decay can be written as

$$\mathrm{sss} \rightarrow \mathrm{uds} + \mathrm{s\bar{u}}.$$

Canceling the common pair of s quarks from each side, we find the net process to be

$$\mathrm{s} \rightarrow \mathrm{u} + \mathrm{d} + \mathrm{\bar{u}},$$

using the quark identities for Λ^0 and K^- given in Tables 52-9 and 52-8. That is, the s quark is transformed into a d quark, and a $\mathrm{u\bar{u}}$ pair is created from the decay energy.

(*c*) Again replacing the particles by their quark content, we can write the reaction as

$$\mathrm{s\bar{u}} + \mathrm{uud} \rightarrow \mathrm{sss} + \mathrm{u\bar{s}} + \mathrm{d\bar{s}},$$

and removing the common quarks of u, d, and s from each side we are left with

$$\mathrm{u\bar{u}} \rightarrow \mathrm{s\bar{s}} + \mathrm{s\bar{s}}.$$

The net process consists of the annihilation of the $\mathrm{u\bar{u}}$ pair and the production of two $\mathrm{s\bar{s}}$ pairs from the reaction energy.

These examples are typical of quark processes: the weak interaction can change one type of quark into another. The strong interaction can create or destroy quark–antiquark pairs, but it cannot change one type of quark into another.

52-5 THE BIG BANG COSMOLOGY

Since the beginning of recorded history, human beings have speculated about the origin and future of the universe, a branch of science now called *cosmology*. Until the 20th century, these speculations were made mostly by philosophers and theologians, because there was no experimental evidence of any sort that would form the basis of any scientific theory. In that century, two major experimental discoveries have pointed the way to a coherent theory that is now accepted by nearly all physicists.

The Expansion of the Universe

The first of the two great discoveries was made by astronomer Edwin Hubble (see Fig. 52-5) in the 1920s. Hubble was studying the wispy objects known previously as *nebulae*. By eventually resolving individual stars in the nebulae, Hubble was able to show that they are galaxies just like our Milky Way, composed of hundreds of billions of stars. More startling, Hubble deduced that the galaxies are moving away from one another and from us, and that the greater their distance from us, the greater is their recessional speed. That is, if d is the distance of the galaxy from Earth (or from any other point of reference in the universe) and v is the speed with which the galaxy appears to be moving away from us, Hubble's law gives

$$v = Hd, \tag{52-2}$$

where H is a proportionality constant known as the *Hubble parameter.*

The Hubble parameter has the dimensions of inverse time. Its value can be learned only by experiment: we must independently deduce the distance of a galaxy from Earth and its speed relative to Earth. The recessional speeds can be measured in a straightforward way using the Doppler shift of the light from the galaxy (see Fig. 39-28), but the distance scale is difficult to determine (in fact, Hubble's early estimates were off by a factor of 10). Figure 52-6 shows an example of more recent data that confirm Hubble's law* and give a range of values of the Hubble parameter. The best set of current data gives a value of the Hubble parameter of

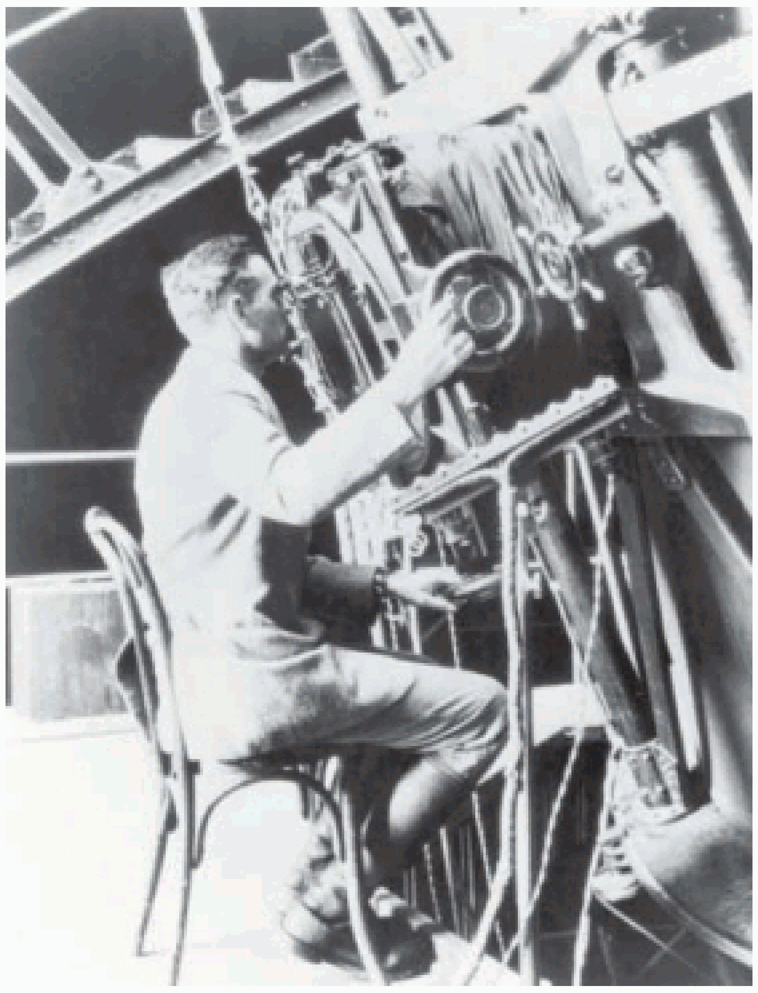


FIGURE 52-5. Edwin Hubble (1889–1953) at the controls of the 100-in. telescope on Mount Wilson, where he did much of the research that led him to propose that the universe is expanding.

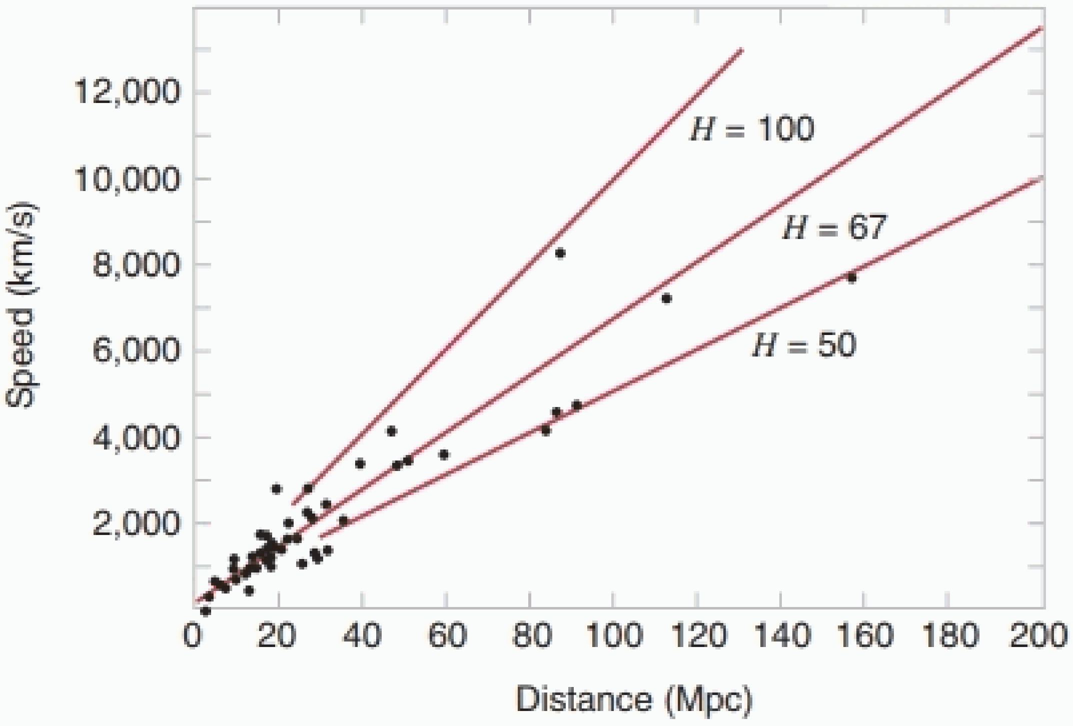


FIGURE 52-6. The relationship between speed and distance for groups and clusters of galaxies. The straight lines show the Hubble relationships for various values of the Hubble parameter H.

$$H = 72\,\frac{\text{km/s}}{\text{Mpc}},$$

where the Mpc (megaparsec) is a commonly used unit of distance on the cosmic scale:

$$\begin{aligned} 1\text{ Mpc} &= 10^6\text{ pc} = 3.26 \times 10^6\text{ light-years} \\ &= 3.084 \times 10^{19}\text{ km}. \end{aligned}$$

Because of uncertainties in the estimates of the cosmic scale of distance, the Hubble parameter is uncertain, with possible values in the range of $65 - 80$ (km/s)/Mpc.

If the universe has been expanding forever at the same rate, then H^{-1} is the age of the universe. Using the accepted value of the Hubble parameter, we would estimate the age of the universe as 14×10^9 y, with the range of uncertainty of H permitting values in the range of $12 - 15 \times 10^9$ y. However, as we shall see later, the expansion of the universe has not been constant, so the true age is less than the currently deduced value of H^{-1}.

The Cosmic Microwave Background Radiation

Although there were other explanations of the expansion of the universe, the one that gained favor was based on the assumption that, if the galaxies are presently rushing apart,

---

*See "The Expansion Rate and Size of the Universe" by Wendy L. Freedman, Scientific American, November 1992. See also http://www.hubbleconstant.com/.

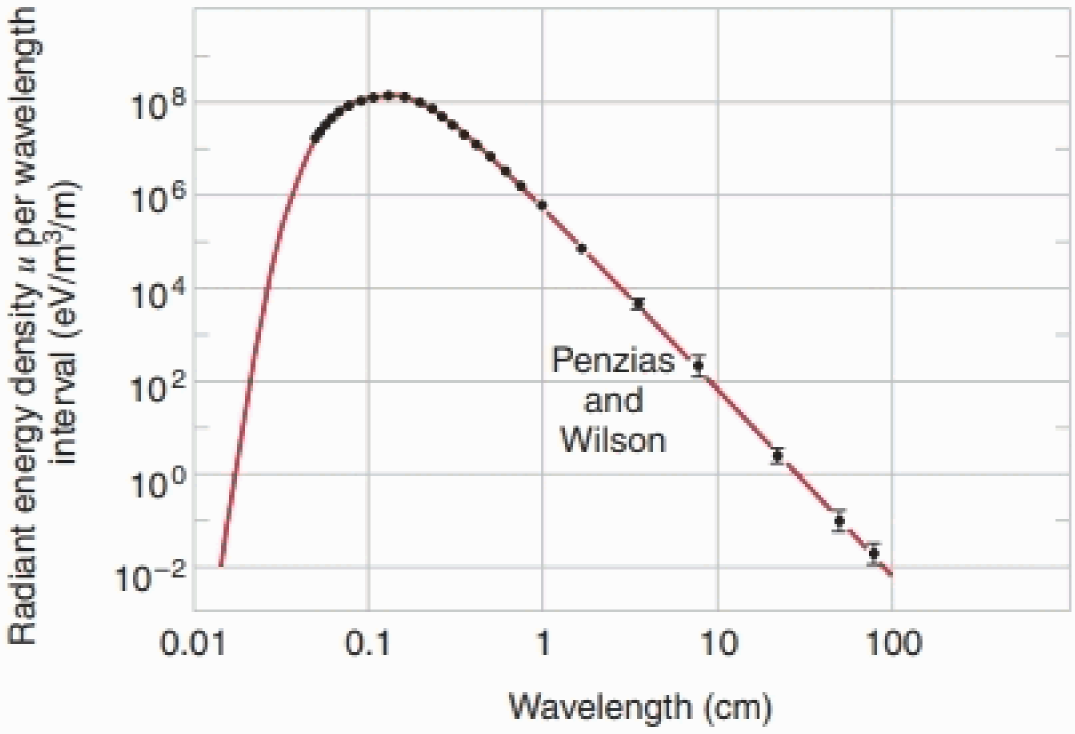


FIGURE 52-8. The spectrum of the cosmic microwave background radiation. The dots represent observations, and the solid line represents the Planck spectrum for the radiant energy corresponding to a temperature of 2.725 K. Note the excellent agreement between the data points and the theoretical curve. The data between 0.05 cm and 1.0 cm come from observations made by the COBE (COsmic Background Explorer) satellite launched in 1989.

they must have been closer together in the distant past. If we run the cosmic clock back far enough, we find that in its early state the universe consisted of unimaginably high densities of matter and radiation. As the universe expanded, both the matter and the radiation cooled; you can think of the wavelengths of the radiant photons being stretched in the expansion. The radiation filled the entire universe in its compact state, and it continues to fill the entire universe in the expansion. We should still find that radiation present today, cooled to the extent that its most intense component is in the microwave region of the electromagnetic spectrum. This is known as the *cosmic microwave background radiation.*

This radiation was discovered in 1965 by Arno Penzias and Robert Wilson of the Bell Laboratories in New Jersey, who were testing a microwave antenna used for satellite communications (see Fig. 52-7). No matter where they pointed their antenna, they found the same annoying background "hiss." Eventually they realized that they were indeed seeing a remnant of the early universe, and they were awarded the 1978 Nobel Prize in physics for their discovery.

The microwave background radiation has a true thermal spectrum of the type we discussed in Section 45-2. Figure 52-8 shows measurements of the intensity of the background radiation at various wavelengths, and you can see how well it is fit by Planck's radiation law with a temperature of 2.725 K. The data points include recent measurements made from a satellite in Earth orbit, thereby eliminating atmospheric absorption.

Measurements of the intensity of the microwave background radiation in various directions show that the radiation has a uniform intensity in all directions; it does not appear to come from any particular source in the sky, but instead fills the entire universe uniformly, as would be expected for radiation that likewise filled the early universe. Recent observations, however, show that there are temperature fluctuations of about 10^{-5} K between different regions of the sky. These results have been interpreted as evidence for the nonuniform distribution of matter in the early universe that led ultimately to the condensation of stars and galaxies.

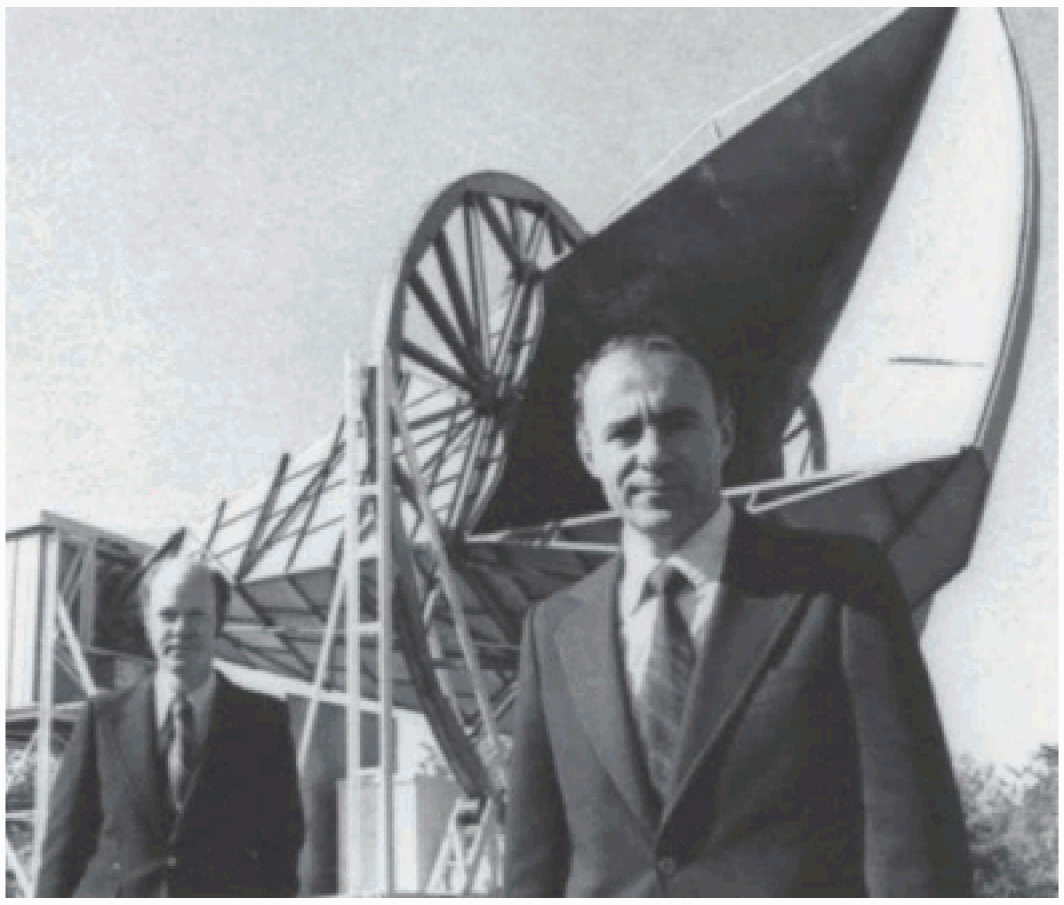


FIGURE 52-7. Arno Penzias (right) and Robert Wilson, standing in front of the large horn antenna with which they first detected the microwave background radiation.

The energy density of the radiation can be found from Planck's radiation law (Eq. 45-6). The number density of these background photons is about 400 per cm^3, and the energy density is about 0.25 eV/cm^3 (roughly corresponding to half the rest energy of an electron per m^3). The mean energy per photon is about 0.00063 eV, which suggests why we are not ordinarily aware of the presence of these photons.

The Big Bang Cosmology

The cosmological theory that is in best agreement with these two experimental findings (the Hubble law and the background radiation) is the *big bang cosmology.* According to this theory, the universe began some 10–15 billion years ago in a state of extreme density and temperature. There were no galaxies or even clumped matter as we now know it; the "stuff" of the universe at early times was a great variety of particles and antiparticles, plus radiation. The density of radiation and matter is related to the temperature of the universe. As the universe expands, it cools (just as any expanding thermodynamic system cools). If we make some reasonable assumptions about the expansion rate, we can find a relationship between the temperature and the time after the formation of the universe:

$$T = \frac{1.5 \times 10^{10}\ \mathrm{s}^{1/2} \cdot \mathrm{K}}{t^{1/2}}, \qquad (52\text{-}3)$$

where the temperature T is in K and the time t is in seconds.

The radiation in the early universe consisted of high-energy photons, whose typical energy can be roughly estimated as kT, where k is the Boltzmann constant and T is the temperature at a particular time t, determined from Eq. 52-3. The dominant processes in the early universe can be represented as

$$\text{photons} \rightarrow \text{particle} + \text{antiparticle}. \qquad (52\text{-}4a)$$

$$\text{particle} + \text{antiparticle} \rightarrow \text{photons} \qquad (52\text{-}4b)$$

Reactions of type 52-4a (called *pair production*) are possible only if the combined energy of the photons on the left side exceeds the total rest energy $2mc^2$ of the particle and antiparticle on the right side. If the temperature is high enough, then the two reactions are each possible; the rates of both reactions are then the same, and there is an equilibrium between the photons and the particles and antiparticles. As the universe expands and cools, the average energy of the photons decreases until at some point, for a specific type of particle, reactions of type 52-4a will no longer be possible. At this point no new particles and antiparticles of this type are being produced, and the equilibrium is upset because reactions of type 52-4b (called *annihilation*) can still proceed. As the particles and antiparticles annihilate one another, their numbers decrease.

Let us consider a specific example in the case of protons. We can represent the reactions as

$$\gamma + \gamma \rightarrow \mathrm{p} + \bar{\mathrm{p}} \quad \text{and} \quad \mathrm{p} + \bar{\mathrm{p}} \rightarrow \gamma + \gamma.$$

The rest energy of the proton is 938 MeV. For the first reaction to occur, the energy on the left side, $2E_\gamma$, must be at least as large as $2m_\mathrm{p}c^2$. If we represent the average photon energy at temperature T as kT, then the corresponding temperature is

$$T = \frac{m_\mathrm{p}c^2}{k} = \frac{938\ \text{MeV}}{8.62 \times 10^{-5}\ \text{eV/K}} = 1.1 \times 10^{13}\ \text{K}.$$

When the temperature of the universe drops below this value, the reaction $\gamma + \gamma \rightarrow \mathrm{p} + \bar{\mathrm{p}}$ will become increasingly less probable, while $\mathrm{p} + \bar{\mathrm{p}} \rightarrow \gamma + \gamma$ will continue to occur. According to Eq. 52-3, this temperature is reached at a time of

$$t = \left(\frac{1.5 \times 10^{10}\ \mathrm{s}^{1/2}\cdot\mathrm{K}}{T}\right)^2 = \left(\frac{1.5 \times 10^{10}\ \mathrm{s}^{1/2}\cdot\mathrm{K}}{1.1. \times 10^{13}\ \mathrm{K}}\right)^2$$
$$= 2 \times 10^{-6}\ \mathrm{s}.$$

That is, at times earlier than about 2 μs, the universe was hot enough to produce proton–antiproton pairs, but at later times only proton-antiproton annihilation occurs.

At very early times, with correspondingly higher temperatures, the radiation may have been hot enough to produce quark–antiquark pairs. If so, at those early times the universe consisted primarily of leptons (and antileptons), quarks (and antiquarks), and photons. The quarks may have come together to form mesons or baryons, but the radiation was sufficiently energetic to dissociate those particles as soon as they formed. Since we do not yet know the details of the interactions between free quarks (or even whether free quarks are allowed to exist), we cannot learn very much about the characteristics of the universe at that time. Instead, we begin the story at a later time, when the universe has cooled sufficiently to allow the quarks and antiquarks to form mesons and baryons.

$t = 10^{-6}$ s ($T = 1.5 \times 10^{13}$ K or $kT = 1300$ MeV). At this time, the universe consists mostly of protons, antiprotons, neutrons, antineutrons, mesons, leptons, antileptons, and photons. The rates of pair production and annihilation (Eqs. 52-4a and 52-4b) are roughly equal, so the number of each kind of particle is about equal to the number of its antiparticle. The number of photons is about equal to the number of protons, which is in turn about equal to the number of electrons.

Because the quarks and antiquarks have formed mesons and baryons, most of the influence of the strong interaction has disappeared by this time. The electromagnetic and weak interactions continue to play an important role. Electromagnetic interactions are represented by processes such as Eqs. 52-4a, b, and weak interactions can occur through such processes as

$$\mathrm{n} + \nu_\mathrm{e} \leftrightarrows \mathrm{p} + \mathrm{e}^- \quad \text{and} \quad \mathrm{p} + \bar{\nu}_\mathrm{e} \leftrightarrows \mathrm{n} + \mathrm{e}^+$$

and similar processes, in which neutrinos are being created and destroyed at the same rate. As long as the leptons have enough energy, the forward and reverse reaction rates are equal, which maintains the balance between the number of charged leptons (e^+ and e^-) and neutrinos. Since these reactions convert neutrons to protons and protons to neutrons with equal ease, the very early universe contained roughly equal numbers of protons and neutrons.

$t = 10^{-2}$ s ($T = 1.5 \times 10^{11}$ K or $kT = 13$ MeV). Now pair production of nucleons (protons and neutrons) no longer occurs. Nucleon–antinucleon annihilation continues to occur; because of a slight excess of matter over antimatter (which we will discuss later), all of the antinucleons disappear, leaving only the small excess of nucleons. The photons are still hot enough to produce electron–positron pairs. At this point the universe consists of protons, neutrons, electrons, positrons, neutrinos, antineutrinos, and photons. The number of protons is roughly equal to the number of neutrons.

$t = 1$ s ($T = 1.5 \times 10^{10}$ K or $kT = 1.3$ MeV). As the universe expands, the neutrinos are cooling along with the photons, and their average energy is also roughly kT. Because the neutron–proton rest energy difference is about 1.8 MeV, the neutrinos no longer have enough energy to convert protons into neutrons; neutrons can be converted to protons, however, such as by neutrino capture ($\mathrm{n} + \nu_\mathrm{e} \rightarrow \mathrm{p} + \mathrm{e}^-$). At this point the neutrinos begin to "decouple" from the matter in the universe; that is, their interactions with matter become much less frequent, and those few remaining interactions have little influence on the properties of the neutrinos (especially their temperature or average energy). From that point until today, the neutrinos continue to fill the universe and to continue their cooling as the universe expands. Today their average energy is very small, about 0.0005 eV, and their density is about 400 per cm^3. As a result of neutrino decoupling the balance between the number of neutrons and protons is upset, and by $t = 1$ s the nucleons consist of about 73% protons and 27% neutrons.

$t = 6$ s ($T = 6 \times 10^9$ K or $kT = 0.5$ MeV). Now the photons have too little energy to produce even electron–positron pairs; electron–positron annihilation has removed all of the positrons and nearly all of the electrons. This further upsets the ability of the transformation between protons and neutrons to go in both directions; as a result, the relative number of protons continues to increase, and by this time we have about 83% protons and 17% neutrons.

At this time the universe consists of a number N protons, $0.2N$ neutrons, N electrons, and a much larger number of photons and neutrinos. At earlier times the number of photons was about equal to the number of nucleons and antinucleons, but the present number of protons represents only the slight excess of protons over antiprotons that remained following annihilation. We can deduce the ratio of photons to nucleons, because it determines the relative amounts of certain light elements such as deuterium (^{2}H) that are formed in the early universe. This number turns out to be about 10^9. That is, for every 1,000,000,001 protons and 1,000,000,000 antiprotons in the early universe, following annihilation there is just one proton and no antiprotons left.

As far as we know, the present universe contains no stars or galaxies made of antimatter. Our universe is entirely made up of the small excess of matter that remained after annihilation. Where did the slight excess of matter over antimatter originate? Evidence for this asymmetry between matter and antimatter is seen in the difference in the decay properties of the neutral K meson K^0 and its antiparticle $\overline{K}^0$. The interaction that causes this asymmetry is not yet well understood, but its effect on the distribution of particles and antiparticles may have established the slight excess of matter during an early era in the evolution of the universe when quarks and leptons were dominant. So far the K^0 is the only system that shows this effect, although it is thought that the neutral B meson B^0 and its antiparticle $\overline{B}^0$ may show a similar effect. A new acceleration and detection facility is currently under construction at the Stanford Linear Accelerator Center to test this prediction. This is another example of the way that results from particle physics have a great impact on developments in cosmology.

This description of the evolution of the universe, illustrated in Fig. 52-9, has taken us from the formation of the universe at the big bang, through hot and turbulent eras dominated by nuclear reactions, to a time of a few seconds when the composition became identical with the particles that now make up our universe. How these particles combined to form the nuclei and atoms that we observe today is discussed in the next section.

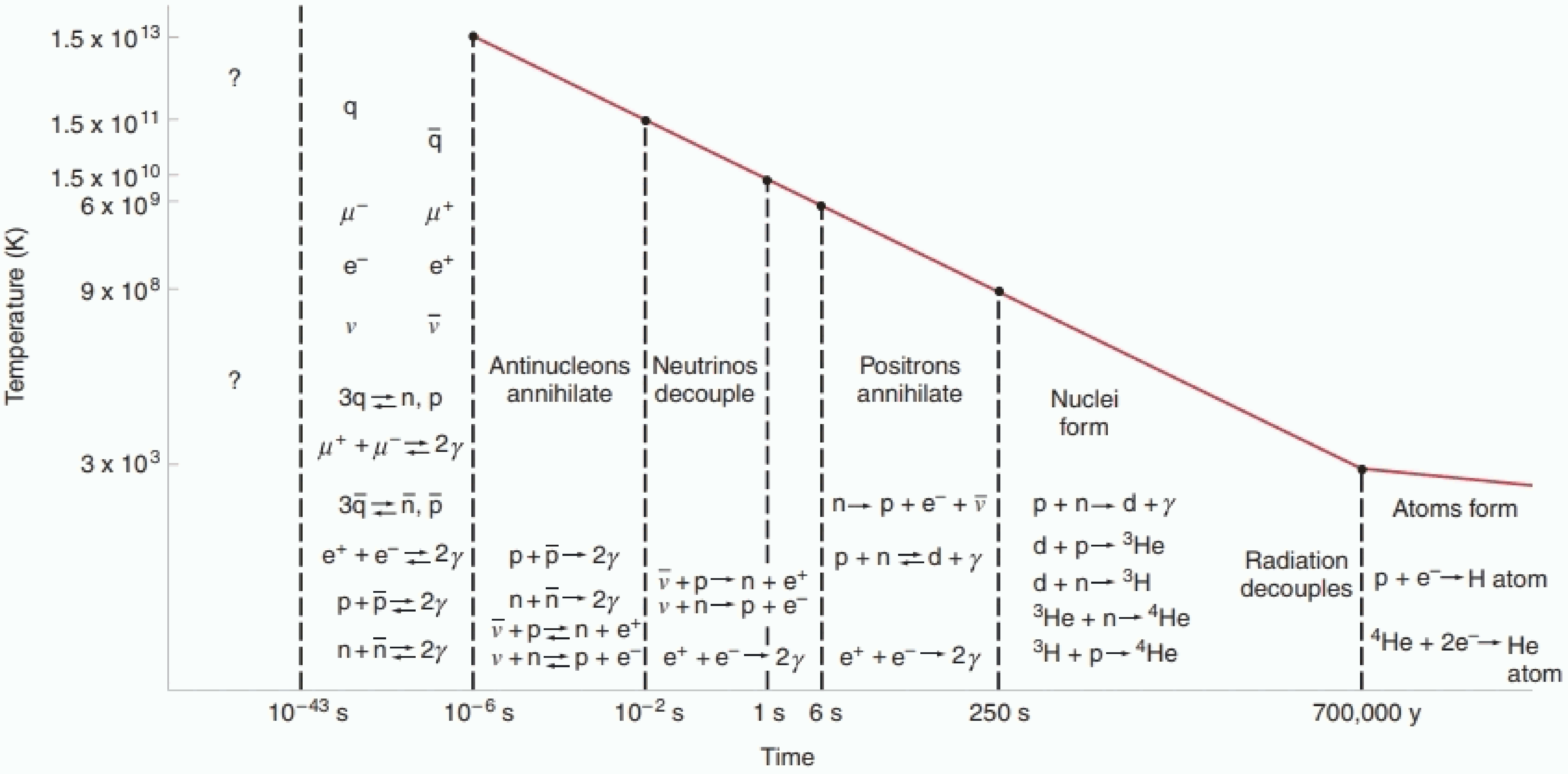


FIGURE 52-9. The evolution of the universe according to the big bang cosmology. The heavy solid line shows the relationship between temperature and time according to Eq. 52-3. The important reactions in each era are shown. (Here q and $\overline{q}$ stand for quark and antiquark, respectively.)

SAMPLE PROBLEM 52-6. When did the universe become too cool to permit the radiation to create $\mu^+\mu^-$ pairs?

Solution The rest energy of the muon is 105.7 MeV. Photons have this mean energy at a temperature determined by

$$T = \frac{m_\mu c^2}{k} = \frac{105.7\text{ MeV}}{8.62 \times 10^{-5}\text{ eV/K}} = 1.23 \times 10^{12}\text{ K}.$$

The corresponding time is found from Eq. 52-3:

$$t = \left(\frac{1.5 \times 10^{10}\text{ s}^{1/2}\cdot\text{K}}{1.23 \times 10^{12}\text{ K}}\right)^2 = 1.5 \times 10^{-4}\text{ s}.$$

52-6 NUCLEOSYNTHESIS

At an age of a few seconds, the universe consisted of protons, neutrons, and electrons. Today, the composition of the universe is mostly hydrogen and helium, with a small abundance of heavier elements. How were the present nuclei and atoms produced from the big bang? The formation of the elements of the present universe is known as *nucleosynthesis.* As we shall see, observing the present abundances of the elements can give us clues about the processes that occurred during the big bang.

Big Bang Nucleosynthesis

The first step in building up complex atoms is the formation of deuterium nuclei (deuterons) from the combination of a proton and a neutron, according to

$$\text{n} + \text{p} \rightarrow \text{d} + \gamma.$$

The binding energy of the deuteron (see Section 50-2) is 2.2 MeV, which is the energy of the γ ray that is given off during the formation. The reverse reaction,

$$\text{d} + \gamma \rightarrow \text{n} + \text{p},$$

can break apart the deuterium nuclei into their constituent protons and neutrons, if the γ ray energy is at least 2.2 MeV.

If the universe is filled with energetic photons, the two reactions will take place at the same rate, and deuterium will be disassociated as quickly as it is formed. However, if the universe is sufficiently old, the photons will not have enough energy to accomplish the disassociation reaction, and deuterium can start to build up.

When we ended our story in the previous section, the universe was about 6 s old, and the mean energy of the radiation was about 0.5 MeV, which is less than what is needed to keep deuterium from forming. However, it must be remembered that the radiation has a Planck distribution of energies (see Fig. 52-10, which was discussed in Section 45-2) and that there are perhaps 10^9 photons for every proton or neutron. There is a high-energy tail in the Planck distribution, which suggests that no matter what the temperature of the radiation, there will always be some photons of energy above 2.2 MeV that can break apart deuterium nuclei. If, on the average, the number of these energetic photons is less than the number of protons and neutrons, deuterium can start to build up.

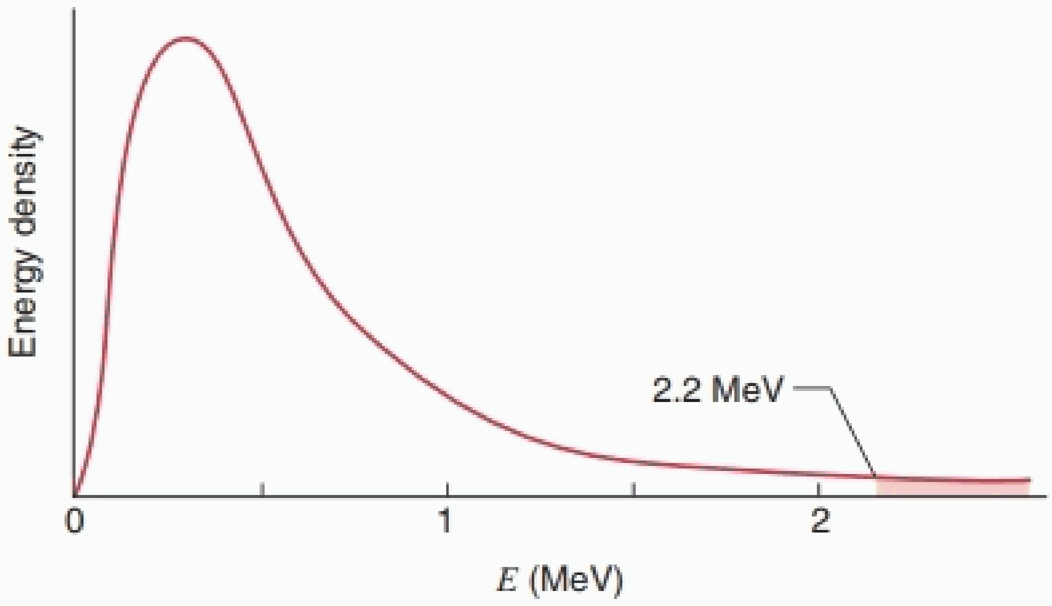


FIGURE 52-10. The energy spectrum of photons at a particular time in the evolution of the universe. Photons with energy above 2.2 MeV, which constitute a tiny fraction of the total number of photons, can dissociate deuterons.

The neutron-to-proton ratio is about 0.2 at this point in the evolution of the universe, and there are roughly 10^9 photons per nucleon, so that the ratio of neutrons to photons is about 0.2×10^{-9}. If the fraction of photons with energies above 2.2 MeV is less than 0.2×10^{-9} of the total number of photons, there will be less than one energetic photon per neutron, and deuterium formation can proceed. From the expression for the Planck distribution (obtained from Eq. 45-6), we find that the fraction of photons of energy greater than 2.2 MeV will be less than 0.2×10^{-9} when the temperature has fallen to 9×10^8 K. Equation 52-3 shows that this temperature occurs at a time of 250 s.

At a time of 250 s, the formation of deuterium nuclei begins. Because the deuterons are less abundant than protons or neutrons, the deuterons will readily react with protons and neutrons, according to the reactions

$$\text{d} + \text{n} \rightarrow {}^3\text{H} + \gamma \quad \text{and} \quad \text{d} + \text{p} \rightarrow {}^3\text{He} + \gamma.$$

Finally, the ^{3}H and ^{3}He will also react with protons and neutrons, as given by

$${}^3\text{H} + \text{p} \rightarrow {}^4\text{He} + \gamma \quad \text{and} \quad {}^3\text{He} + \text{n} \rightarrow {}^4\text{He} + \gamma.$$

For all four of these reactions, the binding energy of the final particle is greater than that of the deuteron. Thus if the radiation is too feeble to prevent the formation of deuterons, it will certainly be too feeble to prevent the succeeding reactions. We can therefore assume that nearly all the deuterons are eventually converted into ^{4}He, so that the end products of this stage of the evolution of the universe are protons and α particles. Because there are no stable nuclei with a mass number of 5, these reactions cannot continue beyond ^{4}He.

To find the relative number of ^{4}He nuclei, we must find the number of available neutrons at $t = 250$ s, when deuterons begin to form. At $t = 6$ s, about 17% of the nucleons are neutrons, but as a result of the radioactive decay of the neutron, some neutrons will be converted into protons between $t = 6$ s and $t = 250$ s. Using the half-life of the neutron (about 11 min), we find that at $t = 250$ s the nucleons will consist of about 12.5% neutrons and 87.5% protons. That is, out of every 10,000 nucleons there will be 1250 neutrons and 8750 protons. These neutrons will combine with 1250 protons to form 625 ^{4}He nuclei, leaving $8750 - 1250 = 7500$ protons. Of the total number of nuclei in the universe at this time, 7.7% are ^{4}He and 92.3% are protons. In terms of mass, the ^{4}He constitutes a fraction of the total mass of the universe given by

$$\frac{4 \times 625}{7500 + 4 \times 625} = 0.25 \text{ or } 25\%.$$

The abundance of ^{4}He in the present universe should equal this value, if we neglect the burning of hydrogen to helium that takes place in stars. The measured helium abundance in a variety of systems, including stars, gaseous nebulae, and planetary nebulae, turns out to be $24 \pm 1\%$, which agrees with our estimate and indicates that our description is certainly reasonable.

The final step in the production of matter in the big bang is the formation of neutral atoms of hydrogen and helium when the protons and ^{4}He nuclei combine with electrons. As in the case of deuteron formation, this cannot occur when there are enough photons in the high-energy tail of the Planck distribution to break apart any neutral atoms that may form. In this case, we want the relative fraction of photons with energies above 13.6 eV (the binding energy of atomic hydrogen) to be less than about 10^{-9}. This occurs for a temperature of about 6000 K, which corresponds to an age of the universe of around 200,000 y. (As the radiation cools, the energy density of the universe becomes less dominated by radiation and more by matter. In this case Eq. 52-3, which assumes a radiation-dominated universe, is not quite correct. Taking this effect into account, the temperature of the universe when hydrogen atoms begin to form is closer to 3000 K, corresponding to an age of around 700,000 y.)

Once neutral atoms have formed, there are essentially no free charged particles left in the universe. This is the time of decoupling of the matter and the radiation field. The universe becomes transparent to the radiation, which can travel long distances without interacting with matter. This radiation, which has been traveling since the decoupling time, is observed today as the microwave background. The expansion of the universe has reduced the radiation temperature by a factor of 1000 since the decoupling time.

The story of the evolution of the universe as described by the big bang cosmology is a remarkable one. It integrates modern experiments in nuclear and particle physics with quantum physics and classical thermodynamics. It yields results that can be tested in the present universe, including the helium abundance, the microwave background radiation, and a small abundance of left-over deuterium that did not get "cooked" into mass-3 nuclei. It is a story that depends in critical ways on the strengths of nuclear or subnuclear forces and on the variety of particles that took part in the early universe. For example, if there were a fourth generation of leptons, the reaction rates of weak-interaction processes would be greater, and more neutrons would be formed, thereby increasing the abundance of ^{4}He. The observed present abundance of ^{4}He is regarded by many cosmologists as limiting the number of generations of leptons to three.

Formation of Heavy Elements

After the decoupling of matter and radiation, the matter (consisting of hydrogen and helium) was subject only to the gravitational force. Recent precise observations of the microwave background have shown that the distribution of matter at the decoupling time was slightly nonuniform. Regions of slightly higher density began to condense into clouds of ever increasing density. As each cloud contracted under its own gravity, its temperature rose until it became hot enough to initiate fusion reactions. This is how first-generation stars formed.

We have seen in Chapter 51 that stars convert hydrogen into helium by means of fusion reactions. After a star has used up its supply of hydrogen and becomes mostly helium, it can again begin to contract, which increases its temperature. (This increase in temperature causes an increased radiation pressure, which causes the radius of the star to increase. The surface area increases more rapidly than the temperature, so that the energy per unit area of the surface actually decreases, and the color of the star goes from bright yellow to red. This is the *red giant* phase of the evolution of the star.) Eventually, the temperature is high enough that the Coulomb barrier between two ^{4}He nuclei can be successfully breached by their thermal motion, and helium fusion can occur. The simple helium fusion reaction

$$^4\text{He} + {}^4\text{He} \rightarrow {}^8\text{Be}$$

does not contribute to the fusion in a star, because ^{8}Be is unstable and breaks apart as rapidly as it forms. Helium fusion requires a third ^{4}He to participate, so that the net reaction is

$$^4\text{He} + {}^4\text{He} + {}^4\text{He} \rightarrow {}^{12}\text{C} + \gamma.$$

Once ^{12}C forms, we can have additional ^{4}He reactions, such as

$$^{12}\text{C} + {}^4\text{He} \rightarrow {}^{16}\text{O} + \gamma,$$

$$^{16}\text{O} + {}^4\text{He} \rightarrow {}^{20}\text{Ne} + \gamma,$$

$$^{20}\text{Ne} + {}^4\text{He} \rightarrow {}^{24}\text{Mg} + \gamma,$$

and so on. These reactions have increasingly high Coulomb barriers and therefore require increasing temperatures.

When the helium fuel is exhausted, contraction sets in again to increase the temperature, so that other reactions can occur, such as carbon burning:

$$^{12}\text{C} + {}^{12}\text{C} \rightarrow {}^{24}\text{Mg} + \gamma.$$

Eventually, these reactions reach the peak of the binding energy curve (Fig. 50-6) at about $A = 56$. Beyond this point, energy is no longer released in fusion reactions.

Figure 52-11 shows the abundance of nuclei in this mass range. The relative abundances support this scenario for producing the elements in fusion reactions. Note that C is more than five orders of magnitude more abundant than Li, Be, and B, which are not produced in these processes. Also note that the even-Z nuclei are, on the average, more than an order of magnitude more abundant than their odd-Z neighbors. The fusion reactions with ^{4}He produce only even-Z products, so the observed higher abundances of these products are consistent with this explanation of their formation.

Note also the last point in Fig. 52-11, which indicates that the *total* abundance of the 50 elements beyond the nuclei in the mass-56 range is less than the abundance of all but one of the individual elements in the region from C to Zn. It certainly appears that most of the matter we know about was produced in fusion processes.

The elements beyond $A = 56$ cannot be produced through fusion reactions in stars. Instead, they are produced through neutron-capture processes. A nucleus in the interior of a star can capture neutrons until its neutron excess is sufficient for it to want to convert an extra neutron to a proton by beta decay: $n \rightarrow p + e^- + \bar{\nu}_e$, thereby increasing the number of protons by one. In this way the atomic number increases one step at a time up to the heaviest nuclei found in nature. For example, the process begins in this way with Fe ($Z = 26$):

$$^{56}\text{Fe} + \text{n} \rightarrow {}^{57}\text{Fe} + \gamma$$

$$^{57}\text{Fe} + \text{n} \rightarrow {}^{58}\text{Fe} + \gamma$$

$$^{58}\text{Fe} + \text{n} \rightarrow {}^{59}\text{Fe} + \gamma$$

Both ^{57}Fe and ^{58}Fe are stable, but ^{59}Fe is radioactive and beta decays with a half-life of 45 days to ^{59}Co ($Z = 27$). The process continues as ^{59}Co can capture a neutron to become ^{60}Co, which is radioactive and beta decays to ^{60}Ni ($Z = 28$). Continuing in this way we can produce all of the heavier elements through the neutron capture–beta decay process.

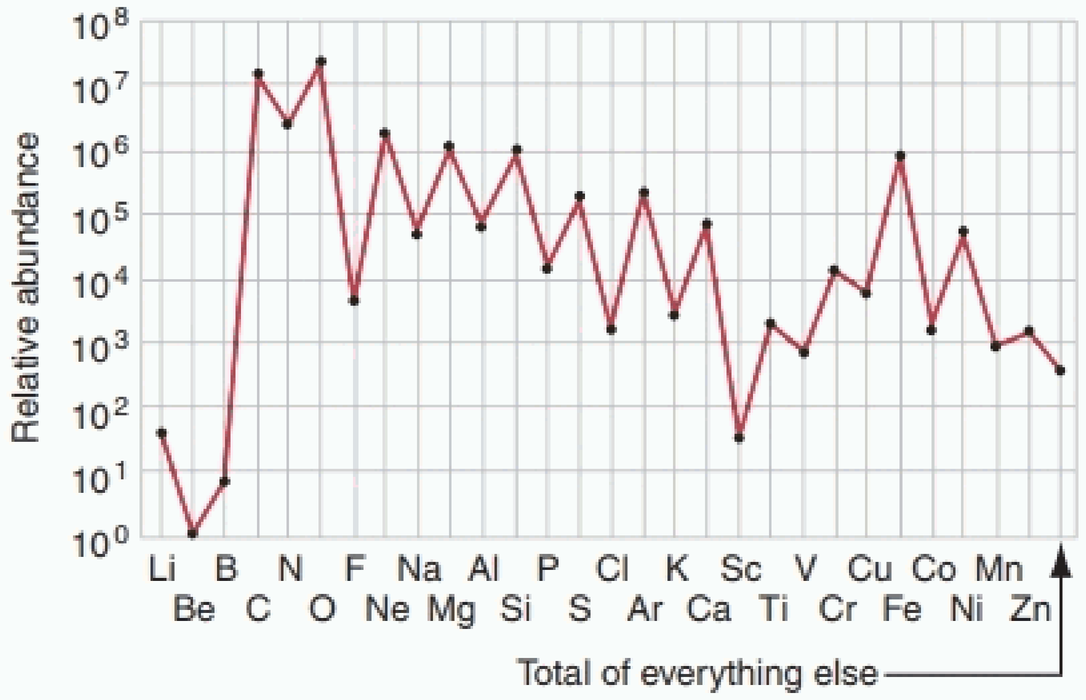


FIGURE 52-11. Relative abundances (by mass) of the elements beyond helium in the solar system.

There is a small density of neutrons in the interiors of stars, and so this process happens slowly during the lifetime of a star. It is therefore known as *s-process* (s for slow) nucleosynthesis. In an explosive supernova, on the other hand, the density of neutrons may be 10^{10}–10^{20} times larger, and the entire process may happen very quickly (seconds or minutes). In this case it is known as *r-process* (r for rapid) nucleosynthesis. Many radioactive nuclei are formed very quickly in the r process and then decay to the stable elements found in nature.

Following either s-process or r-process nucleosynthesis, the resulting elements are distributed into space. The elements beyond mass 56 found on Earth were produced this way in first-generation stars. The planets of our solar system (and in fact we ourselves) are made of the recycled ashes of burnt-out stars.

52-7 THE AGE OF THE UNIVERSE

In Section 50-7, we discussed the use of radioactive dating methods to determine the age of the Earth. By examining the relative amounts of parent and daughter isotopes in certain radioactive decay processes having half-lives in the range $10^8 - 10^9$ y (for example, $^{238}\text{U} \rightarrow {}^{206}\text{Pb}$, $^{87}\text{Rb} \rightarrow {}^{87}\text{Sr}$, and $^{40}\text{K} \rightarrow {}^{40}\text{Ar}$), it has been determined that the age of the oldest rocks on Earth is about 4.5×10^9 y. An identical value is obtained for meteorites and for rocks from the Moon. We can therefore be fairly certain that this value represents the time since the condensation of the solar system.

We know that the universe must be much older than this value, because the solar system formed out of elements that were created in the interiors of stars or in supernovas. The present chemical composition of the solar system was determined during a previous era of nucleosynthesis, which occurred in a previous generation of stars. To find the true age of the universe, we must determine the time interval needed for the elements to be produced.

The total time from the big bang to the present can be divided into four periods: (1) from the big bang until the formation of neutral H and He atoms (t_1); (2) the condensation of galaxies and the formation of first-generation stars (t_2); (3) nucleosynthesis in stars and supernovas, leading to the present chemical elements (t_3); and (4) formation and evolution of the solar system from the debris of earlier stars (t_4). The present age of the universe is the sum of these four terms:

$$t = t_1 + t_2 + t_3 + t_4. \tag{52-5}$$

We know from our discussion of the big bang cosmology that the time t_1 from the big bang until neutral atoms formed is no more than 10^6 y. The time t_2 for galaxies to condense from hydrogen and helium produced in the big bang is not precisely known but has been estimated to be in the range 1–2 × 10^9 y. Since t_4 is known to be 4.5 × 10^9 y, the age of the universe can be determined if we can find the time t_3 associated with nucleosynthesis.

This time must be estimated from the relative abundances of the products that remain at the end of nucleosynthesis. For example, consider the isotopes ^{235}U and ^{238}U, which at present have a relative abundance of 0.72% (see the discussion in connection with the natural fission reactor in Section 51-5). Both ^{235}U and ^{238}U have been decaying during the interval since the formation of the solar system. Their ratio 4.5 × 10^9 years ago is (see Sample Problem 51-4)

$$\begin{aligned} R(0) &= R(t)e^{(\lambda_5 - \lambda_8)t} \\ &= (0.0072)e^{(0.984-0.155)(10^{-9}\,\mathrm{y}^{-1})(4.5\times 10^9\,\mathrm{y})} = 0.30. \end{aligned}$$

During the interval t_3, both isotopes were being formed more or less continuously through the r process, while the relative decay of course also took place. Because of the production of both isotopes during this time, the ratio of their abundances in this period did not change as rapidly as it did during the free decay in the interval t_4; see Fig. 52-12. Evidence from the uranium abundance suggests that t_3 is in the range 4–9 × 10^9 y; the analysis of the abundances of other r-process nuclei gives similar but slightly larger values. The best estimate for t_3 is about 8 × 10^9 y, with a range of 4–12 × 10^9 y.

Combining these results, we have as our estimate for the age of the universe

$$\begin{aligned} t &= t_1 + t_2 + t_3 + t_4 \\ &= 10^6\,\mathrm{y} + 1\text{–}2 \times 10^9\,\mathrm{y} + 8 \times 10^9\,\mathrm{y} + 4.5 \times 10^9\,\mathrm{y} \\ &= 14 \times 10^9\,\mathrm{y}. \end{aligned}$$

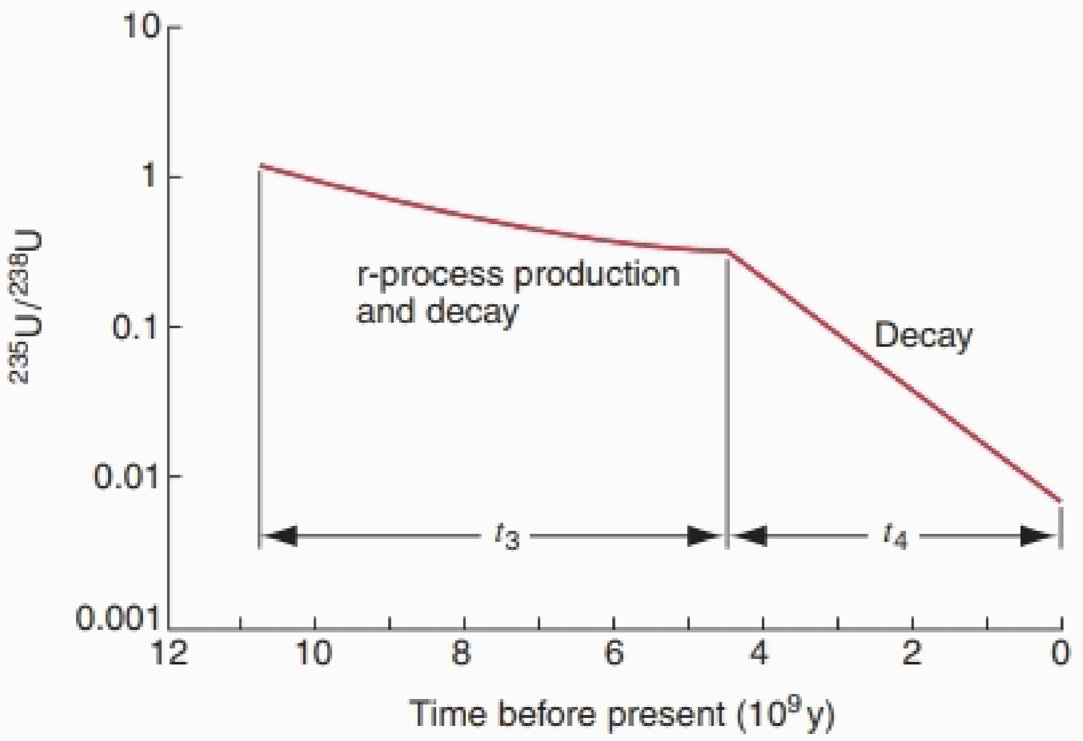


FIGURE 52-12. The change in the ^{235}U/^{238}U ratio with time. During the life of the solar system (the time t_4), the ratio changes due only to the relative decays, eventually reaching the present value of 0.0072. During the interval t_3, production by the r process occurs along with the decay. The duration deduced for the interval t_3 depends on the value that we take for the initial ratio, which must be determined from calculation.

This number is somewhat uncertain as a result of the range of values in the estimate for t_3. Taking this uncertainty into account, we obtain

$$t = 10\text{–}18 \times 10^9\,\mathrm{y}.$$

Consider the enormous amount of physics contained in this simple result. To determine t_1, we used the cumulative knowledge of particle physics, electromagnetism, thermal physics, and atomic and nuclear physics to trace the formation of matter as we know it. The interval t_2 is determined from calculations using thermodynamics and gravitational theory to analyze the condensation of cold matter into hot stars. Our estimate for t_3 is based on our knowledge of r-process and s-process nucleosynthesis based on nuclear physics studies in laboratories on Earth, and the interval t_4 is based on further experiments in nuclear physics and research in geochemistry.

Cosmological Determination of the Age

If we make the rough but not quite correct assumption that the universe has been expanding at the same rate since its formation, then the separation d between typical galaxies should be related to the age of the universe roughly according to

$$d = vt,$$

where v is the (assumed constant) speed of separation. Comparing this result with Eq. 52-2 shows that the age t of the universe is simply the inverse of the Hubble parameter:

$$t = H^{-1}. \tag{52-6}$$

The present best estimate for the Hubble parameter, H = 72 (km/s)/Mpc, gives a value for the age of the universe of

$$t = 14 \times 10^9\,\mathrm{y},$$

in remarkably good agreement with the value obtained from the nucleosynthesis calculation. The range of uncertainty of the Hubble parameter, 65–80 (km/s)/Mpc, gives a corresponding range in ages of

$$t = 12\text{–}15 \times 10^9\,\mathrm{y},$$

which overlaps with the range determined from nucleosynthesis.

Our assumption about the constant separation speed of the galaxies is almost certainly incorrect. The mutual gravitational attraction of the galaxies has been slowing their separation since the big bang, so that at earlier times the speed of separation may have been greater than it is at present. Figure 52-13 shows a representation of a typical intergalactic separation distance as a function of the time. If the "constant-speed" model were valid, the age of the universe would be $t = H^{-1}$. If the speed has been decreasing since the big bang, the deduced age depends on the rate of deceleration. Since humans have not been observing long enough to detect any change in the separation rate, we must

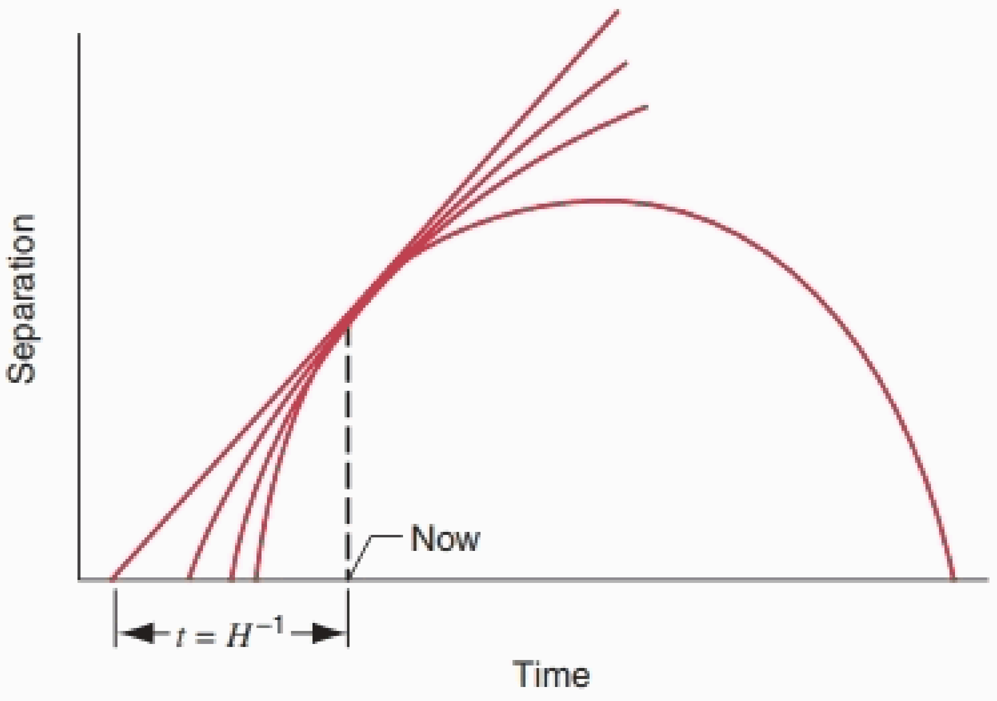


FIGURE 52-13. The dependence of a typical galactic separation distance on time during the evolution of the universe according to different models. If the universe has been expanding at a constant rate (straight line), we can extrapolate backward to zero separation (the big bang) at a time of H^{-1} before the present. If the expansion has been slowing due to the gravitational interaction (a more reasonable scenario), the big bang occurred at a time less than H^{-1} before the present. If the gravitational interaction is strong enough, the expansion may eventually become a contraction.

rely on two indirect methods to determine the deceleration: (1) we can measure the red shifts and thereby deduce the speeds of the most distant (and therefore the oldest) objects we can observe with telescopes, or (2) we can calculate the deceleration based on the gravitational effects of the total amount of matter in the universe.

Somewhat surprisingly, recent measurements of the red shifts of the most distant galaxies suggest that the expansion is *accelerating* rather than decelerating. That is, the expansion rate in the era of these ancient galaxies was slower than we would expect based on extrapolating the current expansion rate. This effect, which has as yet no convincing explanation, has been accounted for through the presence throughout the universe of a "dark energy" that drives the acceleration.

Calculations of the decelerating effect of the matter in the universe are difficult because we don't know how much matter the universe contains. Measurements suggest that the observable matter is not even sufficient to explain the gravitational attraction within galaxies or clusters of galaxies. Astrophysicists have postulated the existence of "dark matter" that is not visible but must be present to produce the gravitational attractions. The amount of this "dark matter" and its form (anything from known or exotic types of elementary particles to burnt-out stars) are uncertain, but it may account for as much as 90% of the mass of the universe.

Various cosmological models have been proposed that give curves in Fig. 52-13 of differing curvatures and therefore different ages of the universe. For example, some of these give an age that is one-half or two-thirds of H^{-1}, or $6-10 \times 10^9$ y. Although we do not know which (if any) of these models is correct, it seems clear that both the nucleosynthesis and cosmological estimates for the age of the universe are consistent with values in the range $10-15 \times 10^9$ y.

It is a source of great frustration for physicists not to be able to view the history of the universe with more certainty, because our ability to look forward is similarly limited. Will the expansion continue forever, or is there enough matter present to reverse the expansion? Figure 52-13 shows several possible outcomes. Perhaps cosmologists of later eras will observe the galaxies rushing together as the universe "heats up" and the galaxies come together, eventually reaching a single point (a "big crunch") that may be followed by another big bang. Or perhaps the expansion continues forever, until the universe is cold and dark. If the solution to this fundamental problem is to be found, it will require vigorous investigations at the forefronts of astrophysics, nuclear physics, and particle physics.

MULTIPLE CHOICE

52-1 Particle Interactions

1. Identify the interaction responsible for the following decays (decay lifetimes are given):

(*a*) $\Delta^* \rightarrow \mathrm{p} + \pi^-$ $(10^{-23}\ \mathrm{s})$; (*b*) $\mathrm{K}^0 \rightarrow \pi^+ + \pi^-$ $(10^{-10}\ \mathrm{s})$; (*c*) $\eta \rightarrow \gamma + \gamma\ (10^{-18}\ \mathrm{s})$.

(A) Strong (B) Electromagnetic
(C) Weak (D) Gravitational

52-2 Families of Particles

2. Given the spin and decay products, to which family does each of these particles belong?

(*a*) Spin $\frac{1}{2}$, decay products are only leptons and/or antileptons.

(*b*) Integer spin, decay products are only leptons and/or antileptons.

(*c*) Spin $\frac{1}{2}$, decay products include one baryon and some mesons.

(*d*) Integer spin, decay products are only mesons.

(*e*) Spin $\frac{1}{2}$, decay products are only mesons.

(A) Lepton (B) Meson
(C) Baryon (D) Not possible

52-3 Conservation Laws

3. Name the conservation law violated in each of the following decays:

(*a*) $\mu^- \rightarrow \mathrm{e}^+ + \nu_\mu + \bar{\nu}_e$; (*b*) $\pi^+ \rightarrow \mathrm{e}^+ + \gamma$; (*c*) $\mu^+ \rightarrow \pi^+ + \bar{\nu}_\mu$.

(A) Electric charge (B) Lepton number
(C) Energy (D) Baryon number

4. Name the conservation law violated in each of the following decays:

(*a*) $\Omega^- \rightarrow \Sigma^- + \pi^0$; (*b*) $\Lambda^0 \rightarrow \pi^+ + \pi^-$;
(*c*) $\Lambda^0 \rightarrow \mathrm{p} + \mathrm{K}^-$.

(A) Energy
(B) Strangeness
(C) Baryon number
(D) Meson number

52-4 The Quark Model

5. If q = quark and $\bar{\mathrm{q}}$ = antiquark, which two of the following combinations might be possible?

(A) $\mathrm{qq\bar{q}}$
(B) $\mathrm{qq\bar{q}\bar{q}}$
(C) $\mathrm{q\bar{q}\bar{q}}$
(D) $\mathrm{qqqq\bar{q}}$
(E) $\mathrm{qqq\bar{q}\bar{q}}$

52-5 The Big Bang Cosmology

6. The energy density of the very early universe ($t < 10^{-6}$ s) was dominated by ________, and the energy density of the present universe is dominated by ________.

(A) radiation, radiation
(B) radiation, matter
(C) matter, radiation
(D) matter, matter

52-6 Nucleosynthesis

52-7 The Age of the Universe

QUESTIONS

1. The ratio of the gravitational force between the electron and the proton in the hydrogen atom to the magnitude of the electromagnetic force of attraction between them is about 10^{-40}. If the gravitational force is so very much weaker than the electromagnetic force, how was it that the gravitational force was discovered first and is so much more apparent to us?

2. What is really meant by an elementary particle? In arriving at an answer, consider such properties as lifetime, mass, size, decays into other particles, fusion to make other particles, and reactions.

3. Why do particle physicists want to accelerate particles to higher and higher energies?

4. Name two particles that have neither mass nor charge. What properties do these particles have?

5. Why do neutrinos leave no tracks in detecting chambers?

6. Neutrinos have no mass (presumably) and travel with the speed of light. How, then, can they carry varying amounts of energy?

7. Do all particles have antiparticles? What about the photon?

8. In the beta decay of an antineutron to an antiproton, which is emitted—a neutrino or an antineutrino?

9. Photons and neutrinos are alike in that they have zero charge, zero mass, and travel with the speed of light. What are the differences between these particles? How would you produce them? How do you detect them?

10. Explain why we say that the π^0 meson is its own antiparticle.

11. An electron cannot decay by disintegrating into two neutrinos. Why not?

12. Why is the electron stable? That is, why does it not decay spontaneously into other particles?

13. A resting electron cannot emit a single gamma-ray photon and disappear. Why? Could a moving electron do so?

14. A neutron is massive enough to decay by the emission of a proton and two neutrinos. Why does it not do so?

15. A positron invariably finds an electron and they annihilate each other. How then can we call a positron a stable particle?

16. What is the mechanism by which two electrons exert forces on each other?

17. Why are particles not grouped into families on the basis of their mass?

18. A particle that responds to the strong force is either a meson or a baryon. You can tell which it is by allowing the particle to decay until only stable end products remain. If there is a proton among these products, the original particle was a baryon. If there is no proton, the original particle was a meson. Explain this classification rule.

19. How many kinds of stable leptons are there? Stable mesons? Stable baryons? In each case, name them.

20. Most particle physics reactions are endothermic rather than exothermic. Why?

21. What is the lightest strongly interacting particle? What is the heaviest particle unaffected by the strong interaction?

22. For each of the following particles, state which of the four basic forces are influential: (*a*) electron, (*b*) neutrinos, (*c*) neutron, (*d*) pion.

23. Just as x rays are used to discover internal imperfections in a metal casting caused by gas bubbles, so cosmic-ray muons have been used in an attempt to discover hidden burial chambers in Egyptian pyramids. Why were muons used?

24. Are strongly interacting particles affected by the weak interaction?

25. Do all weak-interaction decays produce neutrinos?

26. Mesons and baryons are each sensitive to the strong force. In what ways are they different?

27. By comparing Tables 52-3 and 52-7, point out as many similarities between leptons and quarks as you can and also as many differences.

28. What is the experimental evidence for the existence of quarks?

29. We can explain the "ordinary" world around us with two leptons and two quarks. Name them.

30. The neutral pion has the quark structure $\mathrm{u\bar{u}}$ and decays with a mean life of only 8.3×10^{-17} s. The charged pion, on the other hand, has a quark structure of $\mathrm{u\bar{d}}$ and decays with a mean life of 2.6×10^{-8} s. Explain, in terms of their quark structure, why the mean life of the neutral pion should be so much shorter than that of the charged pion. (*Hint:* Think of annihilation.)

31. Do leptons contain quarks? Do mesons? Do photons? Do baryons?

32. The Δ^* baryon can have an electric charge of $+2e$ (see Table 52-5). Based on the quark model, do we expect to find mesons with charge $+2e$? Baryons with charge $-2e$?

33. The Σ^+ baryon decays with a mean life characteristic of the weak interaction (see Table 52-5). It should be able to decay to the Λ^0 baryon by the strong interaction without changing strangeness. Why does it not?

34. Why can we not find the center of the expanding universe? Are we looking for it?

35. Due to the effect of gravity, the rate of expansion of the universe must have decreased in time following the big bang. Show that this implies that the age of the universe is less than $1/H$.

36. It is not possible, using telescopes that are sensitive in any part of the electromagnetic spectrum, to "look back" any farther than about 500,000 y from the big bang. Why?

37. How does one arrive at the conclusion that visible matter may account for only about 10% of the matter in the universe?

38. Are we always looking back in time as we observe a distant galaxy? Does the direction in which we look make a difference?

39. Can you think of any possible explanation for the expanding universe other than a big bang?

EXERCISES

52-1 Particle Interactions

1. (*a*) An electron and a positron are separated by a distance r. Find the ratio of the gravitational force to the electrostatic force between them. What do you conclude from the result concerning the forces acting between particles detected in a bubble chamber or similar detector? (*b*) Repeat for a proton–antiproton pair.

2. Some of the GUTs predict the following possible decay schemes for the proton:

$$\mathrm{p} \rightarrow \mathrm{e}^+ + \gamma,$$
$$\mathrm{p} \rightarrow \mathrm{e}^+ + \pi^0.$$

(*a*) Calculate the Q-values for these decays. (*b*) Show that the decays do not violate the conservation laws of charge, relativistic energy, or linear momentum. The rest energy of a proton is 938.27 MeV, of an electron is 0.511 MeV, and of a neutral pion is 135 MeV.

3. An electron and a proton are placed a distance apart equal to one Bohr radius a_0. Find the radius R of a lead sphere that must be placed directly behind the electron so that the gravitational force on the electron just overcomes the electrostatic attraction between the proton and the electron; see Fig. 52-14. Assume that Newton's law of gravitation holds, and that the density of the sphere equals the density of lead on the Earth.

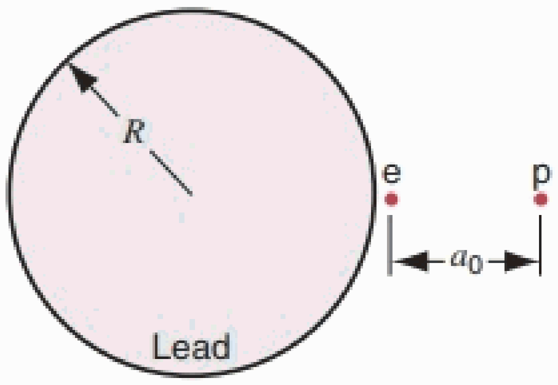


FIGURE 52-14. Exercise 3.

52-2 Families of Particles

4. A neutral pion decays into two gamma rays: $\pi^0 \rightarrow \gamma + \gamma$. Calculate the wavelengths of the gamma rays produced by the decay of a neutral pion at rest.

5. The rest energy of many short-lived particles cannot be measured directly, but must be inferred from the measured momenta and known rest energies of the decay products. Consider the ρ^0 meson, which decays by $\rho^0 \rightarrow \pi^+ + \pi^-$. Calculate the rest energy of the ρ^0 meson given that each of the oppositely directed momenta of the created pions has magnitude 358.3 MeV/c. See Table 52-4 for the rest energies of the pions.

6. Observations of neutrinos emitted by the supernova SN1987a in the Large Magellanic Cloud, see Figs. 52-3 and 52-15, place an upper limit on the rest energy of the electron neutrino of 20 eV. Suppose that the rest energy of the neutrino, rather than being zero, is in fact equal to 20 eV. How much slower than light is a 1.5-MeV neutrino, emitted in a β-decay, moving?

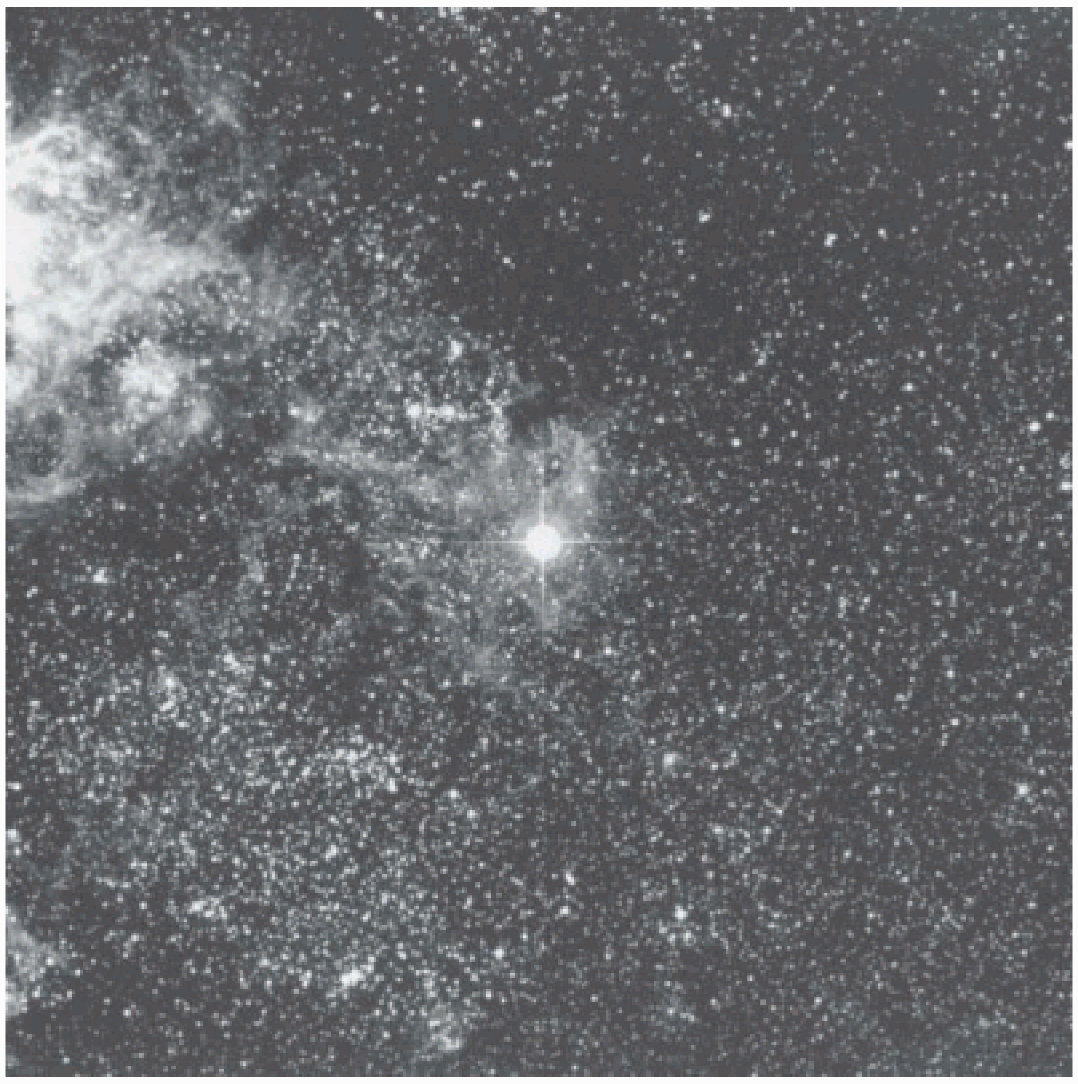


FIGURE 52-15. Exercise 6.

7. Calculate the range of the weak force between two neighboring protons. Assume that the Z^0 boson is the field particle: see Table 52-6.

8. Identify the interaction responsible for each of the following decays: (*a*) $\pi^0 \rightarrow \gamma + \gamma$; (*b*) $\mathrm{K}^+ \rightarrow \mu^+ + \nu_\mu$; (*c*) $\eta' \rightarrow \eta + \pi^+ + \pi^-$; (*d*) $\Sigma^+ \rightarrow \mathrm{p} + \pi^0$.

52-3 Conservation Laws

9. Use the conservation laws to identify the particle labeled x in the following reactions, which proceed by means of the strong interaction: (*a*) $p + p \rightarrow p + \Lambda^0 + x$; (*b*) $p + \bar{p} \rightarrow n + x$; (*c*) $\pi^- + p \rightarrow \Xi^0 + K^0 + x$.

10. The reaction $\pi^+ + p \rightarrow p + p + \bar{n}$ proceeds by the strong interaction. By applying the conservation laws, deduce the charge, baryon number, and strangeness of the antineutron.

11. By examining strangeness, determine which of the following decays or reactions proceed via the strong interaction: (*a*) $K^0 \rightarrow \pi^+ + \pi^-$; (*b*) $\Lambda^0 + p \rightarrow \Sigma^+ + n$; (*c*) $\Lambda^0 \rightarrow p + \pi^-$; (*d*) $K^- + p \rightarrow \Lambda^0 + \pi^0$. See Tables 52-4 and 52-5 for values of S.

12. In addition to the decay mode listed in Table 52-4, give four other possible decays of the K^+ meson that satisfy the conservation laws.

13. Give one possible decay mode for (*a*) $\overline{\Lambda^0}$, (*b*) $\bar{n}$, (*c*) τ^+, (*d*) K^-.

52-4 The Quark Model

14. Make a chart similar to Fig. 52-4 for the 10 spin-$\frac{3}{2}$ baryons listed in Table 52-5.

15. Identify the 10 spin-$\frac{3}{2}$ baryons with the three-quark combinations listed in Table 52-9.

16. Using the up, down, and strange quarks only, construct, if possible, a baryon (*a*) with $Q = +1$ and $S = -2$; (*b*) with $Q = +2$ and $S = 0$.

17. There is no known meson with $Q = +1$ and $S = -1$ or with $Q = -1$ and $S = +1$. Explain why, in terms of the quark model.

18. The quark composition of the proton and the neutron are uud and udd, respectively. What are the quark compositions of (*a*) the antiproton and (*b*) the antineutron?

19. (*a*) List the seven possible quark–antiquark combinations of u, d, c, and s quarks that contain at least one c quark or antiquark. Give the electric charge, strangeness, and charm for each combination. (*b*) Identify the spin-0 mesons in the following table with the quark–antiquark combinations in your list.

Particle	*Antiparticle*	Q	S	C
η_c	η_c	0	0	0
D^+	D^-	+1	0	+1
D^0	$\overline{D^0}$	0	0	+1
D_s^+	D_s^-	+1	+1	+1

20. Analyze the following decays or reactions in terms of the quark content of the particles: (*a*) $\Sigma^- \rightarrow n + \pi^-$; (*b*) $K^0 \rightarrow \pi^+ + \pi^-$; (*c*) $\pi^+ + p \rightarrow \Sigma^+ + K^+$; (*d*) $\gamma + n \rightarrow \pi^- + p$.

52-5 The Big Bang Cosmology

21. By choosing two points on each line of Fig. 52-6 and calculating the slopes, verify the given numerical values of the Hubble parameter.

22. If Hubble's law can be extrapolated to very large distances, at what distance would the recessional speed become equal to the speed of light?

23. Why is the observed wavelength of the 656.3-nm H_α line of hydrogen emitted by a galaxy at a distance of 2.4×10^8 pc?

24. In the laboratory, one of the lines of sodium is emitted at a wavelength of 590.0 nm. When observing the light from a particular galaxy, however, this line is seen at a wavelength of 602.0 nm. Calculate the distance to the galaxy, assuming that Hubble's law holds.

25. (*a*) What is the minimum temperature of the universe necessary for the photons to produce $\pi^+\pi^-$ pairs? (*b*) At what age did the universe have this temperature?

26. (*a*) Estimate the peak wavelength of the microwave background radiation from Fig. 52-8. (*b*) Find the frequency and photon energy corresponding to that wavelength.

27. (*a*) What was the age of the universe when its temperature was 5000 K? In which era did this occur (Fig. 52-9)? (*b*) Estimate the average photon energy at that time. (*c*) If there were 10^9 photons per nucleon at that time, estimate the ratio of the energy of the radiation to the rest energy of the matter.

52-6 Nucleosynthesis

52-7 The Age of the Universe

PROBLEMS

1. A neutral pion has a rest energy of 135 MeV and a mean life of 8.4×10^{-17} s. If it is produced with an initial kinetic energy of 80 MeV and it decays after one mean lifetime, what is the longest possible track that this particle could leave in a bubble chamber? Take relativistic time dilation into account.

2. A positive tau (τ^+, rest energy = 1777 MeV) is moving with 2200 MeV of kinetic energy in a circular path perpendicular to a uniform 1.2-T magnetic field. (*a*) Calculate the momentum of the tau in $\text{kg} \cdot \text{m/s}$. Relativistic effects must be considered. (*b*) Find the radius of the circular path. (*Hint:* See Section 32-3.)

3. The wavelength of the photons at which a radiation field of temperature T radiates most intensely is given by $\lambda_{\text{max}} = (2898\ \mu\text{m} \cdot \text{K})/T$ (see Eq. 45-4). (*a*) Show that the energy E in MeV of such a photon can be computed from

$$E = (4.28 \times 10^{-10}\ \text{MeV/K})T.$$

(*b*) At what minimum temperature can this photon create an electron–positron pair?

4. The recessional speeds of galaxies and quasars at great distances are close to the speed of light, so that the relativistic Doppler shift formula (see Section 39-6) must be used. The redshift is reported as z, where $z = \Delta\lambda/\lambda_0$ is the (fractional) red shift. (*a*) Show that, in terms of z, the recessional speed parameter $\beta = v/c$ is given by

$$\beta = \frac{z^2 + 2z}{z^2 + 2z + 2}.$$

(*b*) One of the most distant quasars detected has $z = 4.43$. Calculate its speed parameter. (*c*) Find the distance to the quasar, assuming that Hubble's law is valid to these distances.

5. Due to the presence everywhere of the microwave radiation background, the minimum possible temperature of a gas in interstellar or intergalactic space is not 0 K but 2.7 K. This implies that a significant fraction of the molecules in space that possess excited states of low excitation energy may, in fact, be in those excited states. Subsequent de-excitation leads to the emission of radiation that could be detected. Consider a (hypothetical) molecule with just one excited state. (*a*) What would the excitation energy have to be in order that 23% of the molecules be in the excited state? (*Hint:* see Section 48-9.) (*b*) Find the wavelength of the photon emitted in a transition to the ground state.

6. Will the universe continue to expand forever? To attack this question, make the (reasonable?) assumption that the recessional speed v of a galaxy a distance r from us is determined only by the matter that lies inside a sphere of radius r centered on us; see Fig. 52-16. If the total mass inside this sphere is M, the escape speed v_e is given by $v_e = \sqrt{2GM/r}$ (see Section 14-6). (*a*) Show that the average density ρ inside the sphere must be at least equal to the value given by

$$\rho = 3H^2/8\pi G$$

to prevent unlimited expansion. (*b*) Evaluate this "critical density" numerically; express your answer in terms of H-atoms/m^3. Measurements of the actual density are difficult and complicated by the presence of dark matter.

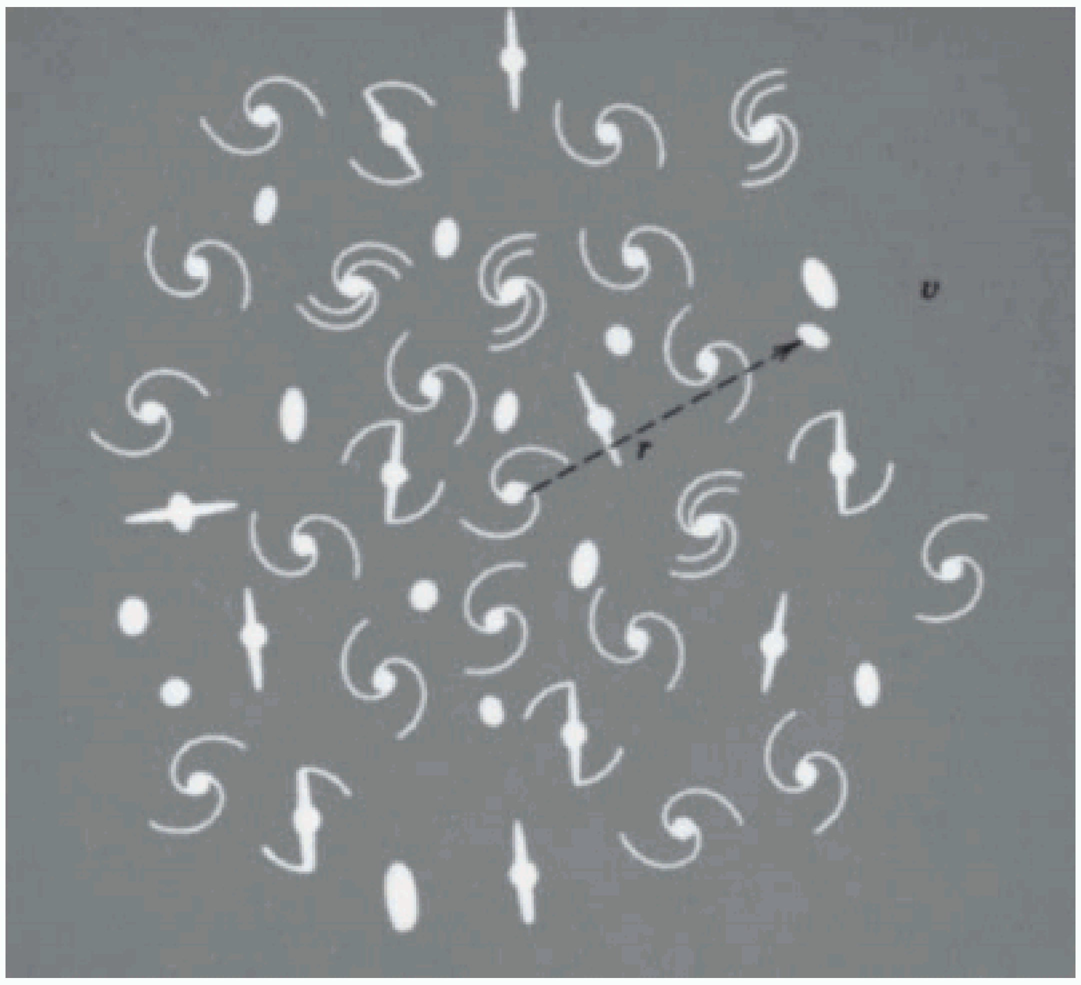


FIGURE 52-16. Problem 6.

7. The existence of dark (i.e., nonluminous) matter in a galaxy (such as our own) can be inferred by determining through observation the variation with distance in the orbital period of revolution of stars about the galactic center. This is then compared with the variation derived on the basis of the distribution of matter as indicated by the luminous material (mostly stars). Any significant deviation implies the existence of dark matter. For example, suppose that the matter (stars, gas, dust) of a particular galaxy, total mass M, is distributed uniformly throughout a sphere of radius R. A star, mass m, is revolving about the center of the galaxy in a circular orbit of radius $r < R$. (*a*) Show that the orbital speed v of the star is given by

$$v = r\sqrt{GM/R^3},$$

and therefore that the period T of revolution is

$$T = 2\pi\sqrt{R^3/GM},$$

independent of r. (*b*) What is the corresponding formula for the orbital period assuming that the mass of the galaxy is strongly concentrated toward the center of the galaxy, so that essentially all of the mass is at distances from the center less than r? These considerations applied to our own Milky Way galaxy indicate that substantial quantities of dark matter are present.

8. Find the temperature at which the fraction of photons with energy greater than 2.2 MeV in the Planck distribution is 0.2×10^{-9}. (Hint: Use Eq. 45-6.)

APPENDIX A

THE INTERNATIONAL SYSTEM OF UNITS (SI)*

The SI Base Units

Quantity	*Name*	*Symbol*	*Definition*
Length	meter	m	". . . the length of the path traveled by light in vacuum in 1/299,792,458 of a second." (1983)
Mass	kilogram	kg	". . . the mass of the international prototype of the kilogram." (1901)
Time	second	s	". . . the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom." (1967)
Electric current	ampere	A	". . . that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per meter of length." (1948)
Thermodynamic temperature	kelvin	K	". . . the fraction 1/273.16 of the thermodynamic temperature of the triple point of water." (1967)
Amount of substance	mole	mol	". . . the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12." (1971)
Luminous intensity	candela	cd	". . . the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of 1/683 watt per steradian." (1979)

---

* Adapted from "Guide for the Use of the International System of Units (SI)," National Bureau of Standards Special Publication 811, 1995 edition. The definitions above were adopted by the General Conference of Weights and Measures, an international body, on the dates shown. In this book we do not use the candela.

Some SI Derived Units

Quantity	*Name of Unit*	*Symbol*	*Equivalent*
Area	square meter	m^2	
Volume	cubic meter	m^3	
Frequency	hertz	Hz	s^{-1}
Mass density (density)	kilogram per cubic meter	kg/m^3	
Speed, velocity	meter per second	m/s	
Angular velocity	radian per second	rad/s	
Acceleration	meter per second squared	m/s^2	
Angular acceleration	radian per second squared	rad/s^2	
Force	newton	N	$kg \cdot m/s^2$
Pressure	pascal	Pa	N/m^2
Work, energy, quantity of heat	joule	J	$N \cdot m$
Power	watt	W	J/s
Quantity of electricity	coulomb	C	$A \cdot s$
Potential difference, electromotive force	volt	V	$N \cdot m/C$
Electric field	volt per meter	V/m	N/C
Electric resistance	ohm	Ω	V/A
Capacitance	farad	F	$A \cdot s/V$
Magnetic flux	weber	Wb	$V \cdot s$
Inductance	henry	H	$V \cdot s/A$
Magnetic field	tesla	T	Wb/m^2, $N/A \cdot m$
Entropy	joule per kelvin	J/K	
Specific heat capacity	joule per kilogram kelvin	$J/(kg \cdot K)$	
Thermal conductivity	watt per meter kelvin	$W/(m \cdot K)$	
Radiant intensity	watt per steradian	W/sr	

The SI Supplementary Units

Quantity	*Name of Unit*	*Symbol*
Plane angle	radian	rad
Solid angle	steradian	sr

APPENDIX B

FUNDAMENTAL PHYSICAL CONSTANTS*

Constant	*Symbol*	*Computational Value*	*Best (1998) Value*	
			Value[a]	*Uncertainty*[b]
Speed of light in a vacuum	c	3.00×10^{8} m/s	2.99792458	exact
Elementary charge	e	1.60×10^{-19} C	1.602176462	0.039
Electric constant (permittivity)	ϵ_0	8.85×10^{-12} F/m	8.85418781762	exact
Magnetic constant (permeability)	μ_0	1.26×10^{-6} H/m	1.25663706143	exact
Electron mass	m_e	9.11×10^{-31} kg	9.10938188	0.079
Electron mass[c]	m_e	5.49×10^{-4} u	5.485799110	0.0021
Proton mass	m_p	1.67×10^{-27} kg	1.67262158	0.079
Proton mass[c]	m_p	1.0073 u	1.00727646688	0.00013
Neutron mass	m_n	1.67×10^{-27} kg	1.67492716	0.079
Neutron mass[c]	m_n	1.0087 u	1.00866491578	0.00054
Electron charge-to-mass ratio	e/m_e	1.76×10^{11} C/kg	1.758820174	0.040
Proton-to-electron mass ratio	m_p/m_e	1840	1836.1526675	0.0021
Planck constant	h	6.63×10^{-34} J·s	6.62606876	0.078
Electron Compton wavelength	λ_e	2.43×10^{-12} m	2.426310215	0.0073
Molar gas constant	R	8.31 J/mol·K	8.314472	1.7
Avogadro constant	N_A	6.02×10^{23} mol^{-1}	6.02214199	0.079
Boltzmann constant	k	1.38×10^{-23} J/K	1.3806503	1.7
Molar volume of ideal gas at STP[d]	V_m	2.24×10^{-2} m^3/mol	2.2413996	1.7
Faraday constant	F	9.65×10^{4} C/mol	9.64853415	0.040
Stefan–Boltzmann constant	σ	5.67×10^{-8} W/m^2·K^4	5.670400	7.0
Rydberg constant	R_∞	1.10×10^{7} m^{-1}	1.0973731568549	0.0000076
Gravitational constant	G	6.67×10^{-11} m^3/s^2·kg	6.673	1500
Bohr radius	a_0	5.29×10^{-11} m	5.291772083	0.0037
Electron magnetic moment	μ_e	9.28×10^{-24} J/T	9.28476362	0.040
Proton magnetic moment	μ_p	1.41×10^{-26} J/T	1.410606633	0.041
Bohr magneton	μ_B	9.27×10^{-24} J/T	9.27400899	0.040
Nuclear magneton	μ_N	5.05×10^{-27} J/T	5.05078317	0.040
Fine structure constant	α	1/137	1/137.03599976	0.0037
Magnetic flux quantum	Φ_0	2.07×10^{-15} Wb	2.067833636	0.039
von Klitzing constant	R_K	25800 Ω	25812.807572	0.0037

[a] Same unit and power of ten as the computational value.

[b] Parts per million.

[c] Mass given in unified atomic mass units, where 1 u = $1.66053873 \times 10^{-27}$ kg.

[d] STP—standard temperature and pressure = 0°C and 1.0 bar.

* *Source:* Peter J. Mohr and Barry N. Taylor, *Journal of Physical and Chemical Reference Data,* vol. 28, no. 6 (1999) and *Reviews of Modern Physics,* vol. 72, no. 2 (2000). See also http://physics.nist.gov/constants.

APPENDIX C

ASTRONOMICAL DATA

The Sun, the Earth, and the Moon

Property	*Sun*[a]	*Earth*	*Moon*
Mass (kg)	1.99×10^{30}	5.98×10^{24}	7.36×10^{22}
Mean radius (m)	6.96×10^{8}	6.37×10^{6}	1.74×10^{6}
Mean density (kg/m^3)	1410	5520	3340
Surface gravity (m/s^2)	274	9.81	1.67
Escape velocity (km/s)	618	11.2	2.38
Period of rotation[c] (d)	26–37[b]	0.997	27.3
Mean orbital radius (km)	2.6×10^{17}[d]	1.50×10^{8}[e]	3.82×10^{5}[f]
Orbital period	2.4×10^{8} y[d]	1.00 y[e]	27.3 d[f]

[a] The Sun radiates energy at the rate of 3.90×10^{26} W; just outside the Earth's atmosphere solar energy is received, assuming normal incidence, at the rate of 1380 W/m^2.

[b] The Sun—a ball of gas—does not rotate as a rigid body. Its rotational period varies between 26 d at the equator and 37 d at the poles.

[c] Measured with respect to the distant stars.

[d] About the galactic center.

[e] About the Sun.

[f] About the Earth.

Some Properties of the Planets

	Mercury	*Venus*	*Earth*	*Mars*	*Jupiter*	*Saturn*	*Uranus*	*Neptune*	*Pluto*
Mean distance from Sun (10^6 km)	57.9	108	150	228	778	1,430	2,870	4,500	5,900
Period of revolution (y)	0.241	0.615	1.00	1.88	11.9	29.5	84.0	165	248
Period of rotation[a] (d)	58.7	243[b]	0.997	1.03	0.409	0.426	0.451[b]	0.658	6.39
Orbital speed (km/s)	47.9	35.0	29.8	24.1	13.1	9.64	6.81	5.43	4.74
Inclination of axis to orbit	$<28°$	$\approx 3°$	23.4°	25.0°	3.08°	26.7°	97.9°	29.6°	57.5°
Inclination of orbit to Earth's orbit	7.00°	3.39°	—	1.85°	1.30°	2.49°	0.77°	1.77°	17.2°
Eccentricity of orbit	0.206	0.0068	0.0167	0.0934	0.0485	0.0556	0.0472	0.0086	0.250
Equatorial diameter (km)	4,880	12,100	12,800	6,790	143,000	120,000	51,800	49,500	2,300
Mass (Earth = 1)	0.0558	0.815	1.000	0.107	318	95.1	14.5	17.2	0.002
Average density (g/cm^3)	5.60	5.20	5.52	3.95	1.31	0.704	1.21	1.67	2.03
Surface gravity[c] (m/s^2)	3.78	8.60	9.78	3.72	22.9	9.05	7.77	11.0	0.03
Escape speed (km/s)	4.3	10.3	11.2	5.0	59.5	35.6	21.2	23.6	1.3
Known satellites	0	0	1	2	16 + rings	19 + rings	15 + rings	8 + rings	1

[a] Measured with respect to the distant stars.

[b] The sense of rotation is opposite to that of the orbital motion.

[c] Measured at the planet's equator.

APPENDIX D

PROPERTIES OF THE ELEMENTS

Element	*Symbol*	*Atomic Number,* *Z*	*Molar Mass* (g/mol)	*Density* (g/cm^3) at 20°C	*Melting Point* (°C)	*Boiling Point* (°C)	*Specific Heat* (J/g·C°) at 25°C
Actinium	Ac	89	(227)	10.1 (calc.)	1051	3200	0.120
Aluminum	Al	13	26.9815	2.699	660	2519	0.897
Americium	Am	95	(243)	13.7	1176	2011	—
Antimony	Sb	51	121.76	6.69	630.6	1587	0.207
Argon	Ar	18	39.948	1.6626×10^{-3}	-189.3	-185.9	0.520
Arsenic	As	33	74.9216	5.72	817 (28 at.)	614 (subl.)	0.329
Astatine	At	85	(210)	—	302	337	—
Barium	Ba	56	137.33	3.5	727	1597	0.204
Berkelium	Bk	97	(247)	14 (est.)	1050	—	—
Beryllium	Be	4	9.0122	1.848	1287	2471	1.83
Bismuth	Bi	83	208.980	9.75	271.4	1564	0.122
Bohrium	Bh	107	(264)	—	—	—	—
Boron	B	5	10.81	2.34	2075	4000	1.03
Bromine	Br	35	79.904	3.12 (liquid)	-7.2	58.8	0.226
Cadmium	Cd	48	112.41	8.65	321.1	767	0.232
Calcium	Ca	20	40.08	1.55	842	1484	0.647
Californium	Cf	98	(251)	—	900 (est.)	—	—
Carbon	C	6	12.011	2.25	3550	—	0.709
Cerium	Ce	58	140.12	6.770	798	3424	0.192
Cesium	Cs	55	132.905	1.873	28.44	671	0.242
Chlorine	Cl	17	35.453	3.214×10^{-3} (0°C)	-101.5	-34.0	0.479
Chromium	Cr	24	51.996	7.19	1907	2671	0.449
Cobalt	Co	27	58.9332	8.85	1495	2927	0.421
Copper	Cu	29	63.54	8.96	1084.6	2562	0.385
Curium	Cm	96	(247)	13.5 (calc.)	1345	—	—
Dubnium	Db	105	(262)	—	—	—	—
Dysprosium	Dy	66	162.50	8.55	1412	2567	0.170
Einsteinium	Es	99	(252)	—	860 (est.)	—	—
Erbium	Er	68	167.26	9.07	1529	2868	0.168
Europium	Eu	63	151.96	5.244	822	1529	0.182
Fermium	Fm	100	(257)	—	1527	—	—
Fluorine	F	9	18.9984	1.696×10^{-3} (0°C)	-219.6	-188.1	0.824
Francium	Fr	87	(223)	—	27	677	—
Gadolinium	Gd	64	157.25	7.90	1313	3273	0.236

(Continued)

Element	*Symbol*	*Atomic Number, Z*	*Molar Mass* (g/mol)	*Density* (g/cm^3) at 20°C	*Melting Point* (°C)	*Boiling Point* (°C)	*Specific Heat* (J/g · C°) at 25°C
Gallium	Ga	31	69.72	5.904	29.76	2204	0.371
Germanium	Ge	32	72.61	5.323	938.3	2833	0.320
Gold	Au	79	196.967	19.3	1064.18	2856	0.129
Hafnium	Hf	72	178.49	13.31	2233	4603	0.144
Hassium	Hs	108	(269)	—	—	—	—
Helium	He	2	4.0026	0.1664×10^{-3}	−272.2	−268.9	5.19
Holmium	Ho	67	164.930	8.79	1474	2700	0.165
Hydrogen	H	1	1.00797	0.08375×10^{-3}	−259.34	−252.87	14.3
Indium	In	49	114.82	7.31	156.6	2072	0.233
Iodine	I	53	126.9044	4.93	113.7	184.4	0.145
Iridium	Ir	77	192.2	22.4	2446	4428	0.131
Iron	Fe	26	55.845	7.87	1538	2861	0.449
Krypton	Kr	36	83.80	3.488×10^{-3}	−157.4	−153.2	0.248
Lanthanum	La	57	138.91	6.145	918	3464	0.195
Lawrencium	Lr	103	(260)	—	—	—	—
Lead	Pb	82	207.19	11.35	327.5	1749	0.129
Lithium	Li	3	6.941	0.534	180.5	1342	3.58
Lutetium	Lu	71	174.97	9.84	1663	3402	0.154
Magnesium	Mg	12	24.305	1.74	650	1090	1.02
Manganese	Mn	25	54.9380	7.43	1244	2061	0.79
Meitnerium	Mt	109	(268)	—	—	—	—
Mendelevium	Md	101	(258)	—	827	—	—
Mercury	Hg	80	200.59	13.55	−38.83	356.7	0.140
Molybdenum	Mo	42	95.94	10.22	2623	4639	0.251
Neodymium	Nd	60	144.24	7.00	1021	3074	0.190
Neon	Ne	10	20.180	0.8387×10^{-3}	−248.6	−246.0	1.03
Neptunium	Np	93	(237)	20.25	644	3902	1.26
Nickel	Ni	28	58.69	8.902	1455	2913	0.444
Niobium	Nb	41	92.906	8.57	2477	4744	0.265
Nitrogen	N	7	14.0067	1.1649×10^{-3}	−210.0	−195.8	1.04
Nobelium	No	102	(259)	—	—	—	—
Osmium	Os	76	190.2	22.57	3033	5012	0.130
Oxygen	O	8	15.9994	1.3318×10^{-3}	−218.8	−183.0	0.918
Palladium	Pd	46	106.4	12.02	1555	2963	0.246
Phosphorus	P	15	30.9738	1.82	44.15	280.5	0.769
Platinum	Pt	78	195.08	21.45	1768	3825	0.133
Plutonium	Pu	94	(244)	19.84	640	3228	0.130
Polonium	Po	84	(209)	9.32	254	962	—
Potassium	K	19	39.098	0.86	63.28	759	0.757
Praseodymium	Pr	59	140.907	6.773	931	3520	0.193
Promethium	Pm	61	(145)	7.264	1042	3000 (est.)	—
Protactinium	Pa	91	(231)	15.4 (calc.)	1572	—	—
Radium	Ra	88	(226)	5.0	700	1140	—
Radon	Rn	86	(222)	9.96×10^{-3} (0°C)	−71	−61.7	0.094
Rhenium	Re	75	186.2	21.02	3186	5596	0.137
Rhodium	Rh	45	102.905	12.41	1964	3695	0.243
Rubidium	Rb	37	85.47	1.53	39.31	688	0.363
Ruthenium	Ru	44	101.07	12.41	2334	4150	0.238
Rutherfordium	Rf	104	(261)	—	—	—	—
Samarium	Sm	62	150.35	7.52	1074	1794	0.197
Scandium	Sc	21	44.956	2.99	1541	2836	0.568
Seaborgium	Sg	106	(266)	—	—	—	—
Selenium	Se	34	78.96	4.79	221	685	0.321
Silicon	Si	14	28.086	2.33	1414	3265	0.705
Silver	Ag	47	107.68	10.49	961.8	2162	0.235
Sodium	Na	11	22.9898	0.971	97.72	883	1.23

(Continued)

Element	*Symbol*	*Atomic Number, Z*	*Molar Mass* (g/mol)	*Density* (g/cm^3) at 20°C	*Melting Point* (°C)	*Boiling Point* (°C)	*Specific Heat* (J/g · C°) at 25°C
Strontium	Sr	38	87.62	2.54	777	1382	0.301
Sulfur	S	16	32.066	2.07	115.2	444.6	0.710
Tantalum	Ta	73	180.948	16.6	3017	5458	0.140
Technetium	Tc	43	(98)	11.5 (calc.)	2157	4265	—
Tellurium	Te	52	127.60	6.24	449.5	988	0.202
Terbium	Tb	65	158.924	8.23	1356	3230	0.182
Thallium	Tl	81	204.38	11.85	304	1473	0.129
Thorium	Th	90	(232)	11.72	1750	4788	0.113
Thulium	Tm	69	168.934	9.32	1545	1950	0.160
Tin	Sn	50	118.71	7.31	231.93	2602	0.228
Titanium	Ti	22	4788	4.54	1668	3287	0.523
Tungsten	W	74	183.85	19.3	3422	5555	0.132
Ununnilium*	Uun	110	(271)	—	—	—	—
Unununium*	Uuu	111	(272)	—	—	—	—
Ununbium*	Uub	112	(277)	—	—	—	—
Ununquadium*	Uuq	114	(285)	—	—	—	—
Ununhexium*	Uuh	116	(289)	—	—	—	—
Ununoctium*	Uuo	118	(293)	—	—	—	—
Uranium	U	92	(238)	18.95	1135	4131	0.116
Vanadium	V	23	50.942	6.11	1910	3407	0.489
Xenon	Xe	54	131.30	5.495×10^{-3}	-111.75	-108.0	0.158
Ytterbium	Yb	70	173.04	6.966	819	1196	0.155
Yttrium	Y	39	88.905	4.469	1522	3345	0.298
Zinc	Zn	30	65.39	7.133	419.53	907	0.388
Zirconium	Zr	40	91.22	6.506	1855	4409	0.278

Values of molar masses correspond to a mole of *atoms* of the element. For diatomic gases (H_2, O_2, N_2, etc.) the mass of a mole of *molecules* is double the tabulated value.

The values in parentheses in the column of molar masses are the mass numbers of the longest-lived isotopes of those elements that are radioactive.

All the physical properties are given for a pressure of one atmosphere except where otherwise specified.

Except for the molar mass, the data for gases are valid only when these are in their usual molecular state, such as H_2, He, O_2, Ne, etc. The specific heats of the gases are the values at constant pressure.

* Temporary names for these elements.

Source: Handbook of Chemistry and Physics, 79th edition (CRC Press, 1998). See also http://www.webelements.com.

APPENDIX E

PERIODIC TABLE OF THE ELEMENTS

	ALKALI METALS (including hydrogen)																		NOBLE GASES
1	1 H																		2 He
2	3 Li	4 Be													5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg													13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca		21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr		39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	57–70	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	88–102	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Uun*	111 Uuu*	112 Uub*	113	114 Uuq*	115	116 Uuh*	117	118 Uuo*

Lanthanide series	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb
Actinide series	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No

* Discovery of these elements has been reported but names for them have not yet been adopted. The symbols shown represent temporary names assigned to the elements. See http://www.webelements.com for recent information on the discovery and properties of the elements.

APPENDIX F

ELEMENTARY PARTICLES

1. THE FUNDAMENTAL PARTICLES

Leptons

Particle	*Symbol*	*Anti-particle*	*Charge (e)*	*Spin ($h/2\pi$)*	*Rest Energy* (MeV)	*Mean Life* (s)	*Typical Decay Products*
Electron	e^-	e^+	-1	1/2	0.511	∞	
Electron neutrino	ν_e	$\bar{\nu}_e$	0	1/2	<0.000015	∞	
Muon	μ^-	μ^+	-1	1/2	105.7	2.2×10^{-6}	$e^- + \bar{\nu}_e + \nu_\mu$
Muon neutrino	ν_μ	$\bar{\nu}_\mu$	0	1/2	<0.19	∞	
Tau	τ^-	τ^+	-1	1/2	1777	2.9×10^{-13}	$\mu^- + \bar{\nu}_\mu + \nu_\tau$
Tau neutrino	ν_τ	$\bar{\nu}_\tau$	0	1/2	<18	∞	

Quarks

Flavor	*Symbol*	*Antiparticle*	*Charge (e)*	*Spin ($h/2\pi$)*	*Rest Energy*[a] (MeV)	*Other Property*
Up	u	$\bar{u}$	$+2/3$	1/2	3	$C = S = T = B = 0$
Down	d	$\bar{d}$	$-1/3$	1/2	6	$C = S = T = B = 0$
Charm	c	$\bar{c}$	$+2/3$	1/2	1300	Charm (C) $= +1$
Strange	s	$\bar{s}$	$-1/3$	1/2	120	Strangeness (S) $= -1$
Top	t	$\bar{t}$	$+2/3$	1/2	174,000	Topness (T) $= +1$
Bottom	b	$\bar{b}$	$-1/3$	1/2	4300	Bottomness (B) $= -1$

Field Particles

Particle	*Symbol*	*Interaction*	*Charge (e)*	*Spin ($h/2\pi$)*	*Rest Energy* (GeV)
Graviton[b]		Gravity	0	2	0
Weak boson	W^+, W^-	Weak	± 1	1	80.4
Weak boson	Z^0	Weak	0	1	91.2
Photon	γ	Electromagnetic	0	1	0
Gluon	g	Strong (color)	0	1	0

2. SOME COMPOSITE PARTICLES

Baryons

Particle	*Symbol*	*Quark Content*	*Anti-particle*	*Charge (e)*	*Spin ($h/2\pi$)*	*Rest Energy* (MeV)	*Mean Life* (s)	*Typical Decay*
Proton	p	uud	$\overline{\mathrm{p}}$	+1	1/2	938	$>10^{33}$	$\pi^0 + e^+$ (?)
Neutron	n	udd	$\overline{\mathrm{n}}$	0	1/2	940	887	$p + e^- + \overline{\nu}_e$
Lambda	Λ^0	uds	$\overline{\Lambda^0}$	0	1/2	1116	2.6×10^{-10}	$p + \pi^-$
Omega	Ω^-	sss	$\overline{\Omega^-}$	−1	3/2	1672	8.2×10^{-11}	$\Lambda^0 + K^-$
Delta	Δ^{++}	uuu	$\overline{\Delta^{++}}$	+2	3/2	1232	5.7×10^{-24}	$p + \pi^+$
Charmed lambda	Λ_c^+	udc	$\overline{\Lambda_c^+}$	+1	1/2	2285	1.9×10^{-13}	$\Lambda^0 + \pi^+$

Mesons

Particle	*Symbol*	*Quark Content*	*Anti-particle*	*Charge (e)*	*Spin ($h/2\pi$)*	*Rest Energy* (MeV)	*Mean Life* (s)	*Typical Decay*
Pion	π^+	$u\overline{d}$	π^-	+1	0	140	2.6×10^{-8}	$\mu^+ + \nu_\mu$
Pion	π^0	$u\overline{u} + d\overline{d}$	π^0	0	0	135	8.4×10^{-17}	$\gamma + \gamma$
Kaon	K^+	$u\overline{s}$	K^-	+1	0	494	1.2×10^{-8}	$\mu^+ + \nu_\mu$
Kaon	K^0	$d\overline{s}$	$\overline{K^0}$	0	0	498	0.9×10^{-10}	$\pi^+ + \pi^-$
Rho	ρ^+	$u\overline{d}$	ρ^-	+1	1	770	4.4×10^{-24}	$\pi^+ + \pi^0$
D-meson	D^+	$c\overline{d}$	D^-	+1	0	1869	1.1×10^{-12}	$K^- + \pi^+ + \pi^+$
Psi	ψ	$c\overline{c}$	ψ	0	1	3097	7.6×10^{-21}	$e^+ + e^-$
B-meson	B^+	$u\overline{b}$	B^-	+1	0	5279	1.6×10^{-12}	$D^- + \pi^+ + \pi^+$
Upsilon	Υ	$b\overline{b}$	Υ	0	1	9460	1.3×10^{-20}	$e^+ + e^-$

[a] The rest energies listed for the quarks are not those associated with free quarks; since no free quarks have yet been observed, measuring their rest energies in the free state has not yet been possible. The tabulated values are effective rest energies corresponding to quarks bound in composite particles.

[b] Particles expected to exist but not yet observed.

Source: "Review of Particle Properties," *European Physical Journal C,* vol. 15 (2000). Also see http://pdg.lbl.gov/.

APPENDIX G

CONVERSION FACTORS

Conversion factors may be read directly from the tables. For example, 1 degree = 2.778×10^{-3} revolutions, so $16.7° = 16.7 \times 2.778 \times 10^{-3}$ rev. The SI quantities are capitalized. Adapted in part from G. Shortley and D. Williams, *Elements of Physics,* Prentice-Hall, 1971.

Plane Angle

	°	′	″	RADIAN	rev
1 degree =	1	60	3600	1.745×10^{-2}	2.778×10^{-3}
1 minute =	1.667×10^{-2}	1	60	2.909×10^{-4}	4.630×10^{-5}
1 second =	2.778×10^{-4}	1.667×10^{-2}	1	4.848×10^{-6}	7.716×10^{-7}
1 RADIAN =	57.30	3438	2.063×10^{5}	1	0.1592
1 revolution =	360	2.16×10^{4}	1.296×10^{6}	6.283	1

Solid Angle

1 sphere = 4π steradians = 12.57 steradians

Length

	cm	METER	km	in.	ft	mi
1 centimeter =	1	10^{-2}	10^{-5}	0.3937	3.281×10^{-2}	6.214×10^{-6}
1 METER =	100	1	10^{-3}	39.37	3.281	6.214×10^{-4}
1 kilometer =	10^{5}	1000	1	3.937×10^{4}	3281	0.6214
1 inch =	2.540	2.540×10^{-2}	2.540×10^{-5}	1	8.333×10^{-2}	1.578×10^{-5}
1 foot =	30.48	0.3048	3.048×10^{-4}	12	1	1.894×10^{-4}
1 mile =	1.609×10^{5}	1609	1.609	6.336×10^{4}	5280	1

1 angström = 10^{-10} m
1 nautical mile = 1852 m
= 1.151 miles = 6076 ft
1 fermi = 10^{-15} m

1 light-year = 9.460×10^{12} km
1 parsec = 3.084×10^{13} km
1 fathom = 6 ft
1 Bohr radius = 5.292×10^{-11} m

1 yard = 3 ft
1 rod = 16.5 ft
1 mil = 10^{-3} in.
1 nm = 10^{-9} m

Area

	METER^2	cm^2	ft^2	in.^{2}
1 SQUARE METER =	1	10^4	10.76	1550
1 square centimeter =	10^{-4}	1	1.076×10^{-3}	0.1550
1 square foot =	9.290×10^{-2}	929.0	1	144
1 square inch =	6.452×10^{-4}	6.452	6.944×10^{-3}	1

1 square mile = 2.788×10^7 ft^2 = 640 acres
1 barn = 10^{-28} m^2

1 acre = 43,560 ft^2
1 hectare = 10^4 m^2 = 2.471 acre

Volume

	METER^3	cm^3	L	ft^3	in.^{3}
1 CUBIC METER =	1	10^6	1000	35.31	6.102×10^4
1 cubic centimeter =	10^{-6}	1	1.000×10^{-3}	3.531×10^{-5}	6.102×10^{-2}
1 liter =	1.000×10^{-3}	1000	1	3.531×10^{-2}	61.02
1 cubic foot =	2.832×10^{-2}	2.832×10^4	28.32	1	1728
1 cubic inch =	1.639×10^{-5}	16.39	1.639×10^{-2}	5.787×10^{-4}	1

1 U.S. fluid gallon = 4 U.S. fluid quarts = 8 U.S. pints = 128 U.S. fluid ounces = 231 in.^{3}
1 British imperial gallon = 277.4 in.^{3} = 1.201 U.S. fluid gallons

Mass

	g	KILOGRAM	slug	u	oz	lb	ton
1 gram =	1	0.001	6.852×10^{-5}	6.022×10^{23}	3.527×10^{-2}	2.205×10^{-3}	1.102×10^{-6}
1 KILOGRAM =	1000	1	6.852×10^{-2}	6.022×10^{26}	35.27	2.205	1.102×10^{-3}
1 slug =	1.459×10^4	14.59	1	8.786×10^{27}	514.8	32.17	1.609×10^{-2}
1 u =	1.661×10^{-24}	1.661×10^{-27}	1.138×10^{-28}	1	5.857×10^{-26}	3.662×10^{-27}	1.830×10^{-30}
1 ounce =	28.35	2.835×10^{-2}	1.943×10^{-3}	1.718×10^{25}	1	6.250×10^{-2}	3.125×10^{-5}
1 pound =	453.6	0.4536	3.108×10^{-2}	2.732×10^{26}	16	1	0.0005
1 ton =	9.072×10^5	907.2	62.16	5.463×10^{29}	3.2×10^4	2000	1

1 metric ton = 1000 kg
Quantities in the colored areas are not mass units but are often used as such. When we write, for example, 1 kg "=" 2.205 lb this means that a kilogram is a *mass* that *weighs* 2.205 pounds under standard condition of gravity ($g = 9.80665$ m/s^2).

Density

	slug/ft^3	KILOGRAM/METER^3	g/cm^3	lb/ft^3	lb/in.^{3}
1 slug per ft^3	1	515.4	0.5154	32.17	1.862×10^{-2}
1 KILOGRAM per METER^3 =	1.940×10^{-3}	1	0.001	6.243×10^{-2}	3.613×10^{-5}
1 gram per cm^3 =	1.940	1000	1	62.43	3.613×10^{-2}
1 pound per ft^3 =	3.108×10^{-2}	16.02	1.602×10^{-2}	1	5.787×10^{-4}
1 pound per in.^{3} =	53.71	2.768×10^4	27.68	1728	1

Quantities in the colored areas are weight densities and, as such, are dimensionally different from mass densities. See note for mass table.

Time

	y	d	h	min	SECOND
1 year =	1	365.25	8.766×10^{3}	5.259×10^{5}	3.156×10^{7}
1 day =	2.738×10^{-3}	1	24	1440	8.640×10^{4}
1 hour =	1.141×10^{-4}	4.167×10^{-2}	1	60	3600
1 minute =	1.901×10^{-6}	6.944×10^{-4}	1.667×10^{-2}	1	60
1 SECOND =	3.169×10^{-8}	1.157×10^{-5}	2.778×10^{-4}	1.667×10^{-2}	1

Speed

	ft/s	km/h	METER/SECOND	mi/h	cm/s
1 foot per second =	1	1.097	0.3048	0.6818	30.48
1 kilometer per hour =	0.9113	1	0.2778	0.6214	27.78
1 METER per SECOND =	3.281	3.6	1	2.237	100
1 mile per hour =	1.467	1.609	0.4470	1	44.70
1 centimeter per second =	3.281×10^{-2}	3.6×10^{-2}	0.01	2.237×10^{-2}	1

1 knot = 1 nautical mi/h = 1.688 ft/s   1 mi/min = 88.00 ft/s = 60.00 mi/h

Force

	dyne	NEWTON	lb	pdl	gf	kgf
1 dyne =	1	10^{-5}	2.248×10^{-6}	7.233×10^{-5}	1.020×10^{-3}	1.020×10^{-6}
1 NEWTON =	10^{5}	1	0.2248	7.233	102.0	0.1020
1 pound =	4.448×10^{5}	4.448	1	32.17	453.6	0.4536
1 poundal =	1.383×10^{4}	0.1383	3.108×10^{-2}	1	14.10	1.410×10^{-2}
1 gram-force =	980.7	9.807×10^{-3}	2.205×10^{-3}	7.093×10^{-2}	1	0.001
1 kilogram-force =	9.807×10^{5}	9.807	2.205	70.93	1000	1

Quantities in the colored areas are not force units but are often used as such. For instance, if we write 1 gram-force "=" 980.7 dynes, we mean that a gram-mass experiences a force of 980.7 dynes under standard conditions of gravity ($g = 9.80665\ \mathrm{m/s^2}$).

Energy, Work, Heat

	Btu	erg	ft · lb	hp · h	JOULE	cal	kW · h	eV	MeV	kg	u
1 British thermal unit =	1	1.055×10^{10}	777.9	3.929×10^{-4}	1055	252.0	2.930×10^{-4}	6.585×10^{21}	6.585×10^{15}	1.174×10^{-14}	7.070×10^{12}
1 erg =	9.481×10^{-11}	1	7.376×10^{-8}	3.725×10^{-14}	10^{-7}	2.389×10^{-8}	2.778×10^{-14}	6.242×10^{11}	6.242×10^{5}	1.113×10^{-24}	670.2
1 foot-pound =	1.285×10^{-3}	1.356×10^{7}	1	5.051×10^{-7}	1.356	0.3238	3.766×10^{-7}	8.464×10^{18}	8.464×10^{12}	1.509×10^{-17}	9.037×10^{9}
1 horsepower-hour =	2545	2.685×10^{13}	1.980×10^{6}	1	2.685×10^{6}	6.413×10^{5}	0.7457	1.676×10^{25}	1.676×10^{19}	2.988×10^{-11}	1.799×10^{16}
1 JOULE =	9.481×10^{-4}	10^{7}	0.7376	3.725×10^{-7}	1	0.2389	2.778×10^{-7}	6.242×10^{18}	6.242×10^{12}	1.113×10^{-17}	6.702×10^{9}
1 calorie =	3.969×10^{-3}	4.186×10^{7}	3.088	1.560×10^{-6}	4.186	1	1.163×10^{-6}	2.613×10^{19}	2.613×10^{13}	4.660×10^{-17}	2.806×10^{10}
1 kilowatt-hour =	3413	3.6×10^{13}	2.655×10^{6}	1.341	3.6×10^{6}	8.600×10^{5}	1	2.247×10^{25}	2.247×10^{19}	4.007×10^{-11}	2.413×10^{16}
1 electron volt =	1.519×10^{-22}	1.602×10^{-12}	1.182×10^{-19}	5.967×10^{-26}	1.602×10^{-19}	3.827×10^{-20}	4.450×10^{-26}	1	10^{-6}	1.783×10^{-36}	1.074×10^{-9}
1 million electron volts =	1.519×10^{-16}	1.602×10^{-6}	1.182×10^{-13}	5.967×10^{-20}	1.602×10^{-13}	3.827×10^{-14}	4.450×10^{-20}	10^{6}	1	1.783×10^{-30}	1.074×10^{-3}
1 kilogram =	8.521×10^{13}	8.987×10^{23}	6.629×10^{16}	3.348×10^{10}	8.987×10^{16}	2.146×10^{16}	2.497×10^{10}	5.610×10^{35}	5.610×10^{29}	1	6.022×10^{26}
1 unified atomic mass unit =	1.415×10^{-13}	1.492×10^{-3}	1.101×10^{-10}	5.559×10^{-17}	1.492×10^{-10}	3.564×10^{-11}	4.146×10^{-17}	9.32×10^{8}	932.0	1.661×10^{-27}	1

Quantities in the colored areas are not properly energy units but are included for convenience. They arise from the relativistic mass–energy equivalence formula $E = mc^2$ and represent the energy equivalent of a mass of one kilogram or one unified atomic mass unit (u).

Pressure

	atm	dyne/cm^2	inch of water	cm Hg	PASCAL	lb/in.^{2}	lb/ft^2
1 atmosphere =	1	1.013×10^{6}	406.8	76	1.013×10^{5}	14.70	2116
1 dyne per cm^2 =	9.869×10^{-7}	1	4.015×10^{-4}	7.501×10^{-5}	0.1	1.405×10^{-5}	2.089×10^{-3}
1 inch of water° at 4°C =	2.458×10^{-3}	2491	1	0.1868	249.1	3.613×10^{-2}	5.202
1 centimeter of mercury[a] at 0°C =	1.316×10^{-2}	1.333×10^{4}	5.353	1	1333	0.1934	27.85
1 PASCAL =	9.869×10^{-6}	10	4.015×10^{-3}	7.501×10^{-4}	1	1.450×10^{-4}	2.089×10^{-2}
1 pound per in.^{2} =	6.805×10^{-2}	6.895×10^{4}	27.68	5.171	6.895×10^{3}	1	144
1 pound per ft^2 =	4.725×10^{-4}	478.8	0.1922	3.591×10^{-2}	47.88	6.944×10^{-3}	1

[a] Where the acceleration of gravity has the standard value 9.80665 m/s^2.

1 bar = 10^6 dyne/cm^2 = 0.1 MPa   1 millibar = 10^3 dyne/cm^2 = 10^2 Pa   1 torr = 1 millimeter of mercury

Power

	Btu/h	ft · lb/s	hp	cal/s	kW	WATT
1 British thermal unit per hour =	1	0.2161	3.929×10^{-4}	6.998×10^{-2}	2.930×10^{-4}	0.2930
1 foot-pound per second =	4.628	1	1.818×10^{-3}	0.3239	1.356×10^{-3}	1.356
1 horsepower =	2545	550	1	178.1	0.7457	745.7
1 calorie per second =	14.29	3.088	5.615×10^{-3}	1	4.186×10^{-3}	4.186
1 kilowatt =	3413	737.6	1.341	238.9	1	1000
1 WATT =	3.413	0.7376	1.341×10^{-3}	0.2389	0.001	1

Magnetic Flux

	maxwell	WEBER
1 maxwell =	1	10^{-8}
1 WEBER =	10^{8}	1

Magnetic Field

	gauss	TESLA	milligauss
1 gauss =	1	10^{-4}	1000
1 TESLA =	10^{4}	1	10^{7}
1 milligauss =	0.001	10^{-7}	1

1 tesla = 1 weber/meter^2

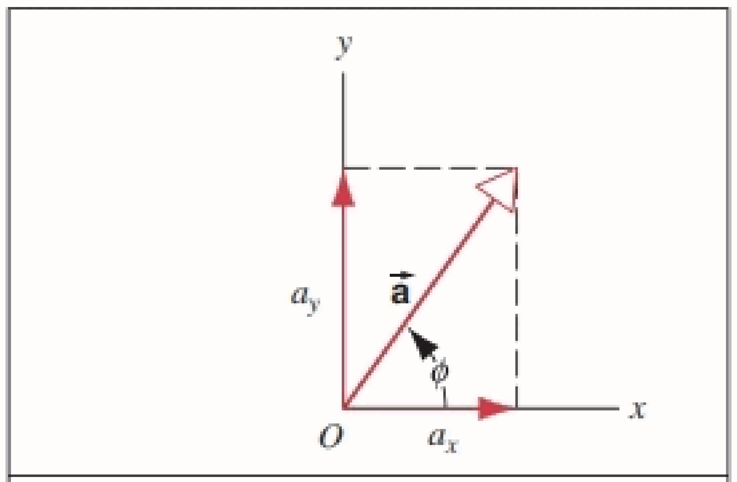
APPENDIX H

VECTORS

H-1 COMPONENTS OF VECTORS

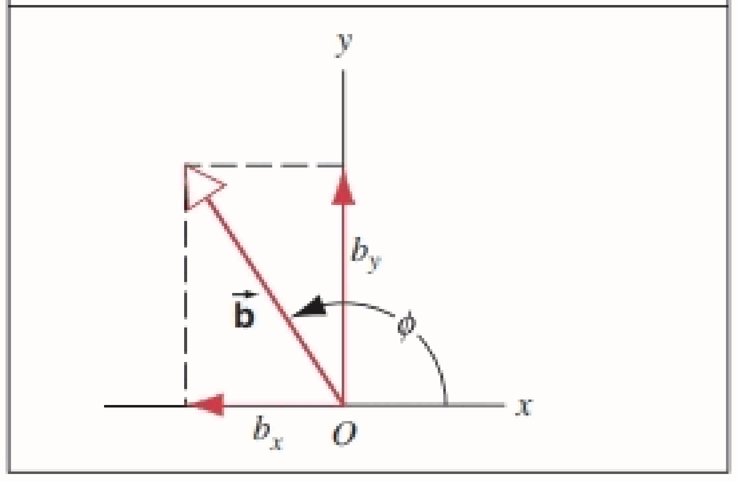
$a_x = a\cos\phi \qquad a_y = a\sin\phi$

$a = \sqrt{a_x^2 + a_y^2} \qquad \tan\phi = a_y/a_x$


y
a_y
$\vec{a}$
ϕ
O
a_x
x

$b_x = b\cos\phi\ (<0)$

$b_y = b\sin\phi\ (>0)$


y
b_y
$\vec{b}$
ϕ
b_x
O
x

$a_x = a\sin\theta\cos\phi$

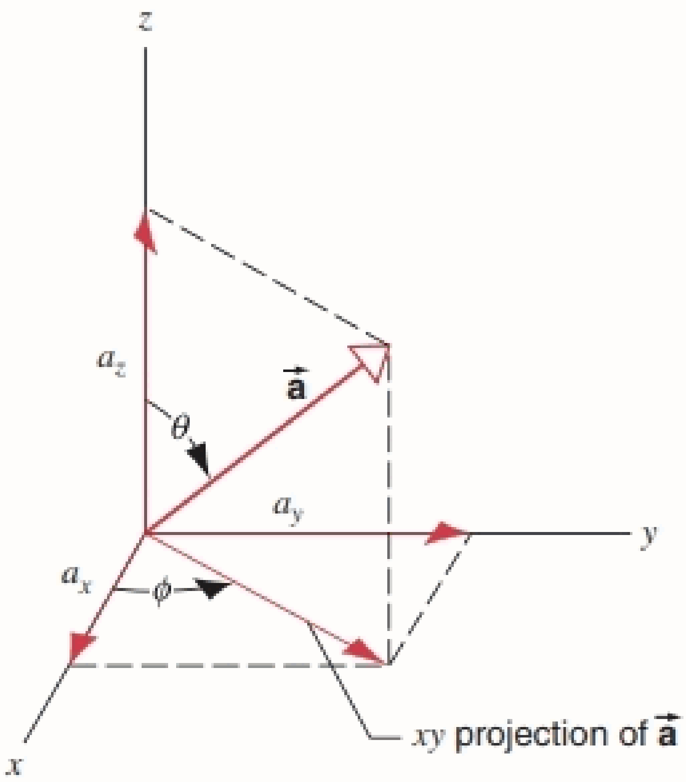
$a_y = a\sin\theta\sin\phi$

$a_z = a\cos\theta$

$a = \sqrt{a_x^2 + a_y^2 + a_z^2}$

$\tan\phi = a_y/a_x$

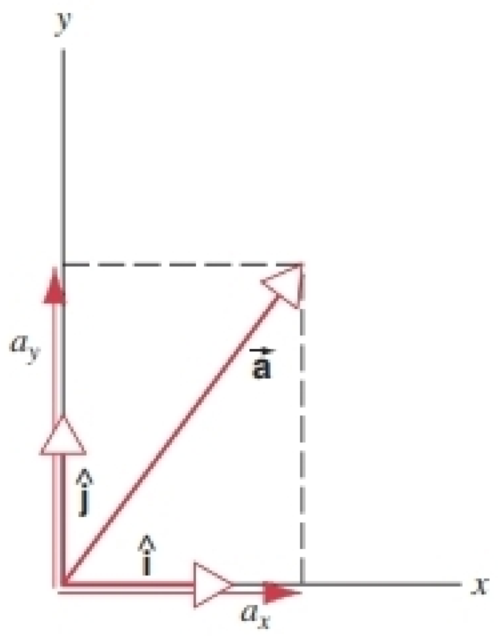
$\cos\theta = a_z/a$


z
a_z
$\vec{a}$
θ
a_y
y
a_x
ϕ
x
xy projection of $\vec{a}$

H-2 UNIT VECTORS

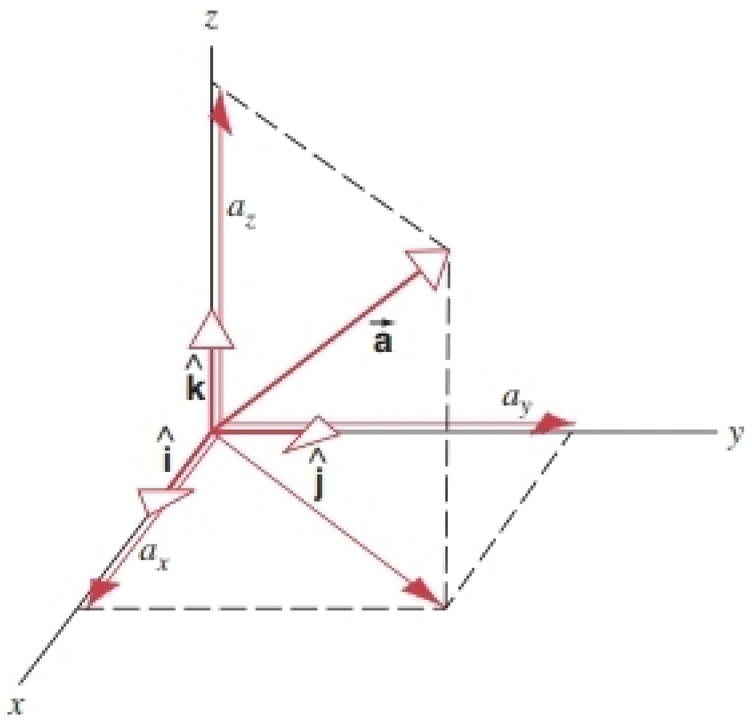
Two-dimensional Cartesian:

$\vec{\mathbf{a}} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}}$


y
a_y
$\vec{\mathbf{a}}$
$\hat{\mathbf{j}}$
$\hat{\mathbf{i}}$
a_x
x

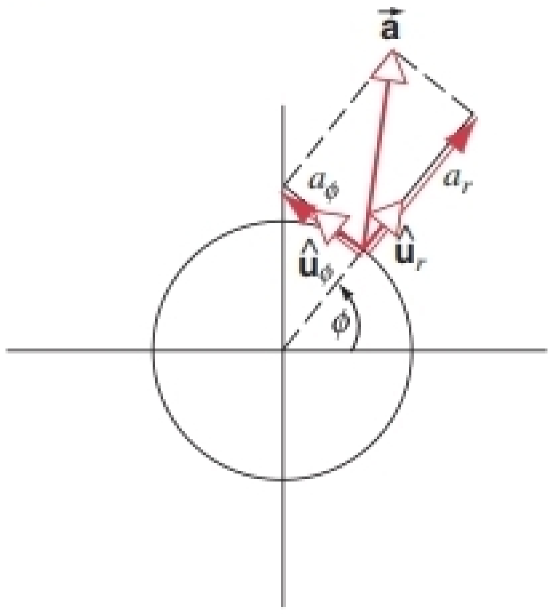
Three-dimensional Cartesian:

$\vec{\mathbf{a}} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}$


z
a_z
$\vec{\mathbf{a}}$
$\hat{\mathbf{k}}$
a_y
y
$\hat{\mathbf{i}}$
$\hat{\mathbf{j}}$
a_x
x

Two-dimensional polar:

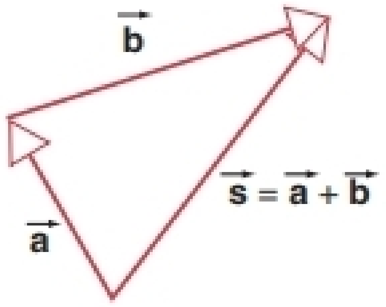
$\vec{\mathbf{a}} = a_r\hat{\mathbf{u}}_r + a_\phi\hat{\mathbf{u}}_\phi$


$\vec{\mathbf{a}}$
a_ϕ
a_r
$\hat{\mathbf{u}}_\phi$
$\hat{\mathbf{u}}_r$
ϕ

H-3 ADDING VECTORS

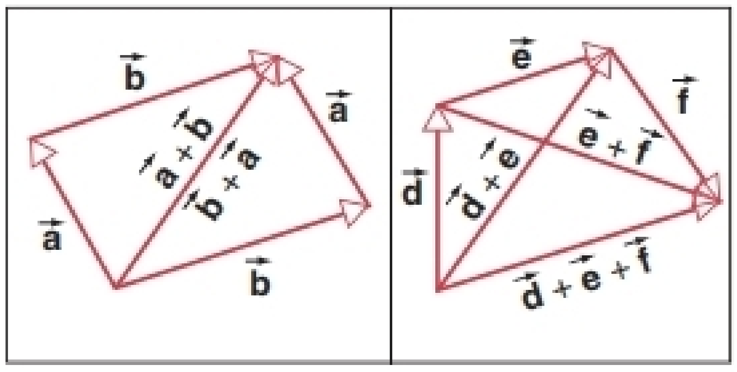
$\vec{\mathbf{s}} = \vec{\mathbf{a}} + \vec{\mathbf{b}}$

$s_x = a_x + b_x \qquad s_y = a_y + b_y$


$\vec{\mathbf{b}}$
$\vec{\mathbf{s}} = \vec{\mathbf{a}} + \vec{\mathbf{b}}$
$\vec{\mathbf{a}}$

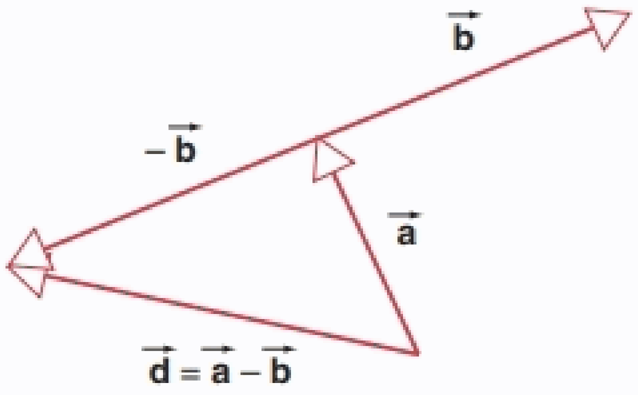
$\vec{\mathbf{a}} + \vec{\mathbf{b}} = \vec{\mathbf{b}} + \vec{\mathbf{a}}$ (commutative law)

$\vec{\mathbf{d}} + (\vec{\mathbf{e}} + \vec{\mathbf{f}}) = (\vec{\mathbf{d}} + \vec{\mathbf{e}}) + \vec{\mathbf{f}}$ (associative law)


$\vec{\mathbf{b}}$
$\vec{\mathbf{a}}$
$\vec{\mathbf{a}} + \vec{\mathbf{b}}$
$\vec{\mathbf{b}} + \vec{\mathbf{a}}$
$\vec{\mathbf{a}}$
$\vec{\mathbf{b}}$
$\vec{\mathbf{e}}$
$\vec{\mathbf{f}}$
$\vec{\mathbf{d}}$
$\vec{\mathbf{d}} + \vec{\mathbf{e}}$
$\vec{\mathbf{e}} + \vec{\mathbf{f}}$
$\vec{\mathbf{d}} + \vec{\mathbf{e}} + \vec{\mathbf{f}}$

$$\vec{d} = \vec{a} - \vec{b} = \vec{a} + (-\vec{b})$$

$$d_x = a_x - b_x \qquad d_y = a_y - b_y$$



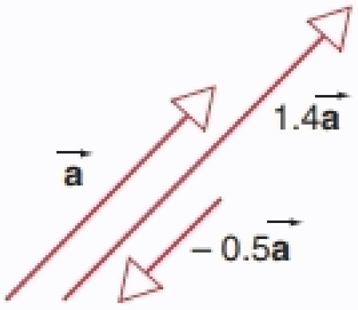
H-4 MULTIPLICATION OF VECTORS

Multiplication of a vector by a scalar:

$$\vec{b} = c\vec{a}$$

$$b_x = ca_x \qquad b_y = ca_y$$

$$b = |c|\,a$$



Dot product (or scalar product) of two vectors:

$$\vec{a} \cdot \vec{b} = ab\cos\phi = a(b\cos\phi) = b(a\cos\phi)$$

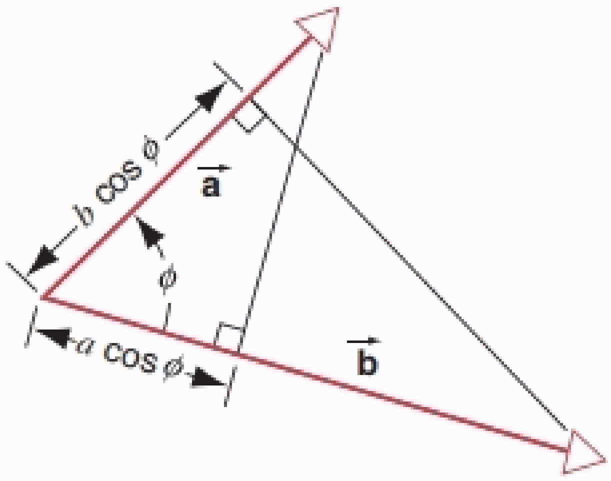
$$\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$$

$$\hat{i} \cdot \hat{i} = \hat{j} \cdot \hat{j} = \hat{k} \cdot \hat{k} = 1$$

$$\hat{i} \cdot \hat{j} = \hat{i} \cdot \hat{k} = \hat{j} \cdot \hat{k} = 0$$

$$\vec{a} \cdot \vec{b} = a_x b_x + a_y b_y + a_z b_z$$

$$\vec{a} \cdot \vec{a} = a^2 = a_x^2 + a_y^2 + a_z^2$$

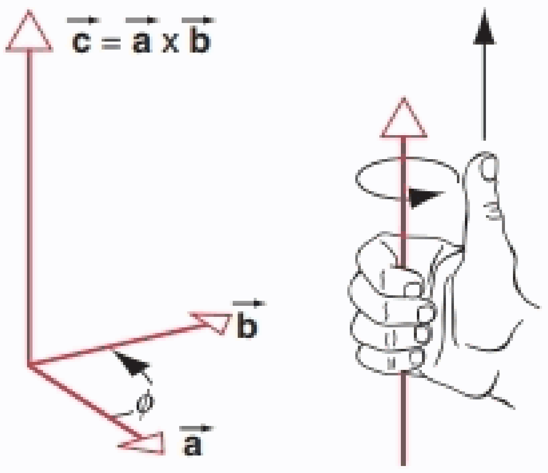


Cross product (or vector product) of two vectors:

$$\vec{c} = \vec{a} \times \vec{b}$$

$$|\vec{c}| = |\vec{a} \times \vec{b}| = ab\sin\phi$$

Direction of $\vec{c}$ is perpendicular to the plane of $\vec{a}$ and $\vec{b}$, determined by the right-hand rule.



$$\vec{b} \times \vec{a} = -\vec{a} \times \vec{b}$$

$$\hat{i} \times \hat{i} = \hat{j} \times \hat{j} = \hat{k} \times \hat{k} = 0$$

$$\hat{i} \times \hat{j} = \hat{k} \qquad \hat{j} \times \hat{k} = \hat{i} \qquad \hat{k} \times \hat{i} = \hat{j}$$

$$\vec{a} \times \vec{b} = (a_y b_z - a_z b_y)\hat{i} + (a_z b_x - a_x b_z)\hat{j} + (a_x b_y - a_y b_x)\hat{k} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}$$

$$\vec{a} \times (\vec{b} + \vec{c}) = (\vec{a} \times \vec{b}) + (\vec{a} \times \vec{c})$$

$$(s\vec{a}) \times \vec{b} = \vec{a} \times (s\vec{b}) = s(\vec{a} \times \vec{b}) \qquad (s = \text{a scalar}).$$

$$\vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{b} \cdot (\vec{c} \times \vec{a}) = \vec{c} \cdot (\vec{a} \times \vec{b})$$

$$\vec{a} \times (\vec{b} \times \vec{c}) = (\vec{a} \cdot \vec{c})\vec{b} - (\vec{a} \cdot \vec{b})\vec{c}$$

APPENDIX I

MATHEMATICAL FORMULAS

Geometry

Circle of radius r: circumference $= 2\pi r$; area $= \pi r^2$.
Sphere of radius r: area $= 4\pi r^2$; volume $= \frac{4}{3}\pi r^3$.
Right circular cylinder of radius r and height h:
area $= 2\pi r^2 + 2\pi r h$; volume $= \pi r^2 h$.
Triangle of base a and altitude h: area $= \frac{1}{2}ah$.

Quadratic Formula

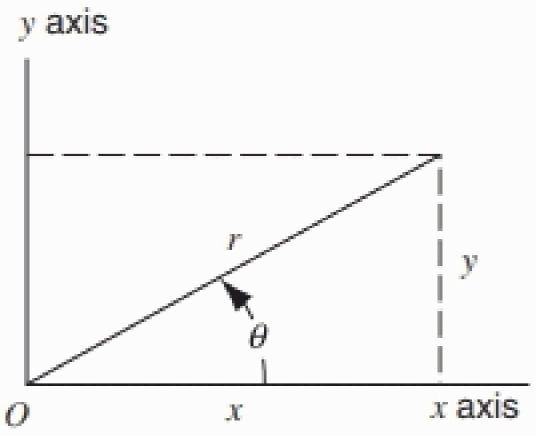
If $ax^2 + bx + c = 0$, then $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

Trigonometric Functions of Angle θ

$$\sin\theta = \frac{y}{r} \qquad \cos\theta = \frac{x}{r}$$

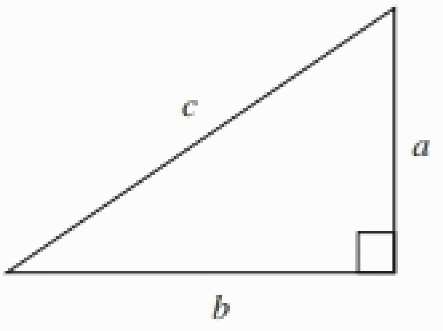
$$\tan\theta = \frac{y}{x} \qquad \cot\theta = \frac{x}{y}$$

$$\sec\theta = \frac{r}{x} \qquad \csc\theta = \frac{r}{y}$$



Pythagorean Theorem

$a^2 + b^2 = c^2$



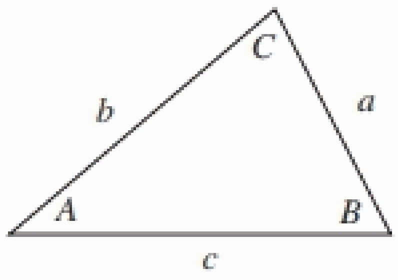
Triangles

Angles A, B, C
Opposite sides a, b, c

$A + B + C = 180°$

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$$

$c^2 = a^2 + b^2 - 2ab\cos C$



Mathematical Signs and Symbols

$=$ equals
$\approx$ equals approximately
$\neq$ is not equal to
$\equiv$ is identical to, is defined as
$>$ is greater than ($\gg$ is much greater than)
$<$ is less than ($\ll$ is much less than)
$\geq$ is greater than or equal to (or, is no less than)
$\leq$ is less than or equal to (or, is no more than)
$\pm$ plus or minus ($\sqrt{4} = \pm 2$)
$\propto$ is proportional to
Σ the sum of
$\bar{x}$ the average value of x (also x_{av})

Trigonometric Identities

$\sin(90° - \theta) = \cos\theta$
$\cos(90° - \theta) = \sin\theta$
$\sin\theta/\cos\theta = \tan\theta$
$\sin^2\theta + \cos^2\theta = 1 \qquad \sec^2\theta - \tan^2\theta = 1 \qquad \csc^2\theta - \cot^2\theta = 1$
$\sin 2\theta = 2\sin\theta\cos\theta$
$\cos 2\theta = \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta$
$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$
$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$
$\tan(\alpha \pm \beta) = \dfrac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$
$\sin\alpha \pm \sin\beta = 2\sin\frac{1}{2}(\alpha \pm \beta)\cos\frac{1}{2}(\alpha \mp \beta)$

Binomial Expansion

$$(1 \pm x)^n = 1 \pm \frac{nx}{1!} + \frac{n(n-1)x^2}{2!} + \cdots \quad (x^2 < 1)$$

$$(1 \pm x)^{-n} = 1 \mp \frac{nx}{1!} + \frac{n(n+1)x^2}{2!} \mp \cdots \quad (x^2 < 1)$$

Exponential Expansion

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

Logarithmic Expansion

$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots \quad (|x| < 1)$

Trigonometric Expansions (θ in radians)

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots$$

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots$$

$$\tan\theta = \theta + \frac{\theta^3}{3} + \frac{2\theta^5}{15} + \cdots$$

Derivatives and Integrals

In what follows, the letters u and v stand for any functions of x, and a and m are constants. To each of the indefinite integrals should be added an arbitrary constant of integration. The *Handbook of Chemistry and Physics* (CRC Press Inc.) gives a more extensive tabulation.

1. $\dfrac{dx}{dx} = 1$
2. $\dfrac{d}{dx}(au) = a\dfrac{du}{dx}$
3. $\dfrac{d}{dx}(u+v) = \dfrac{du}{dx} + \dfrac{dv}{dx}$
4. $\dfrac{d}{dx}x^m = mx^{m-1}$
5. $\dfrac{d}{dx}\ln x = \dfrac{1}{x}$
6. $\dfrac{d}{dx}(uv) = u\dfrac{dv}{dx} + v\dfrac{du}{dx}$
7. $\dfrac{d}{dx}e^x = e^x$
8. $\dfrac{d}{dx}\sin x = \cos x$
9. $\dfrac{d}{dx}\cos x = -\sin x$
10. $\dfrac{d}{dx}\tan x = \sec^2 x$
11. $\dfrac{d}{dx}\cot x = -\csc^2 x$
12. $\dfrac{d}{dx}\sec x = \tan x \sec x$
13. $\dfrac{d}{dx}\csc x = -\cot x \csc x$
14. $\dfrac{d}{dx}e^u = e^u\dfrac{du}{dx}$
15. $\dfrac{d}{dx}\sin u = \cos u\dfrac{du}{dx}$
16. $\dfrac{d}{dx}\cos u = -\sin u\dfrac{du}{dx}$

1. $\int dx = x$
2. $\int au\,dx = a\int u\,dx$
3. $\int (u+v)\,dx = \int u\,dx + \int v\,dx$
4. $\int x^m\,dx = \dfrac{x^{m+1}}{m+1} \quad (m \neq -1)$
5. $\int \dfrac{dx}{x} = \ln|x|$
6. $\int u\dfrac{dv}{dx}\,dx = uv - \int v\dfrac{du}{dx}\,dx$
7. $\int e^x\,dx = e^x$
8. $\int \sin x\,dx = -\cos x$
9. $\int \cos x\,dx = \sin x$
10. $\int \tan x\,dx = -\ln\cos x$
11. $\int \sin^2 x\,dx = \frac{1}{2}x - \frac{1}{4}\sin 2x$
12. $\int \cos^2 x\,dx = \frac{1}{2}x + \frac{1}{4}\sin 2x$
13. $\int e^{-ax}\,dx = -\dfrac{1}{a}e^{-ax}$
14. $\int xe^{-ax}\,dx = -\dfrac{1}{a^2}(ax+1)e^{-ax}$
15. $\int x^2e^{-ax}\,dx = -\dfrac{1}{a^3}(a^2x^2 + 2ax + 2)e^{-ax}$
16. $\int x^ne^{-ax}\,dx = \dfrac{n!}{a^{n+1}}$
17. $\int_0^\infty x^{2n}e^{-ax^2}\,dx = \dfrac{1\cdot 3\cdot 5\cdots(2n-1)}{2^{n+1}a^n}\sqrt{\dfrac{\pi}{a}}$
18. $\int \dfrac{dx}{\sqrt{(x^2 \pm a^2)^3}} = \dfrac{\pm x}{a^2\sqrt{x^2 \pm a^2}}$

APPENDIX J

NOBEL PRIZES IN PHYSICS*

1901	Wilhelm Konrad Röntgen	1845–1923	for the discovery of x-rays
1902	Hendrik Antoon Lorentz	1853–1928	for their researches into the influence of magnetism on radiation phenomena
	Pieter Zeeman	1865–1943	
1903	Antoine Henri Becquerel	1852–1908	for his discovery of spontaneous radioactivity
	Pierre Curie	1859–1906	for their joint researches on the radiation phenomena discovered by Professor Henri Becquerel
	Marie Sklowdowska-Curie	1867–1934	
1904	Lord Rayleigh (John William Strutt)	1842–1919	for his investigations of the densities of the most important gases and for his discovery of argon
1905	Philipp Eduard Anton von Lenard	1862–1947	for his work on cathode rays
1906	Joseph John Thomson	1856–1940	for his theoretical and experimental investigations on the conduction of electricity by gases
1907	Albert Abraham Michelson	1852–1931	for his optical precision instruments and metrological investigations carried out with their aid
1908	Gabriel Lippmann	1845–1921	for his method of reproducing colors photographically based on the phenomena of interference
1909	Guglielmo Marconi	1874–1937	for their contributions to the development of wireless telegraphy
	Carl Ferdinand Braun	1850–1918	
1910	Johannes Diderik van der Waals	1837–1923	for his work on the equation of state for gases and liquids
1911	Wilhelm Wien	1864–1928	for his discoveries regarding the laws governing the radiation of heat
1912	Nils Gustaf Dalén	1869–1937	for his invention of automatic regulators for use in conjunction with gas accumulators for illuminating lighthouses and buoys
1913	Heike Kamerlingh Onnes	1853–1926	for his investigations of the properties of matter at low temperatures, which led, *inter alia,* to the production of liquid helium
1914	Max von Laue	1879–1960	for his discovery of the diffraction of Röntgen rays by crystals
1915	William Henry Bragg	1862–1942	for their services in the analysis of crystal structure by means of X rays
	William Lawrence Bragg	1890–1971	
1917	Charles Glover Barkla	1877–1944	for his discovery of the characteristic X rays of the elements
1918	Max Planck	1858–1947	for his discovery of energy quanta
1919	Johannes Stark	1874–1957	for his discovery of the Doppler effect in canal rays and the splitting of spectral lines in electric fields
1920	Charles-Édouard Guillaume	1861–1938	for the service he has rendered to precision measurements in Physics by his discovery of anomalies in nickel steel alloys
1921	Albert Einstein	1879–1955	for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect
1922	Neils Bohr	1885–1962	for the investigation of the structure of atoms, and of the radiation emanating from them
1923	Robert Andrews Millikan	1868–1953	for his work on the elementary charge of electricity and on the photoelectric effect

---

* See *Nobel Lectures, Physics,* 1901–1970, Elsevier Publishing Company, for biographies of the awardees and for lectures given by them on receiving the prize. For more information, see http://www.nobel.se/physics/laureates/index.html.

1924	Karl Manne Georg Siegbahn	1886–1978	for his discoveries and research in the field of X-ray spectroscopy
1925	James Franck	1882–1964	for their discovery of the laws governing the impact of an electron on an atom
	Gustav Hertz	1887–1975	
1926	Jean Baptiste Perrin	1870–1942	for his work on the discontinuous structure of matter, and especially for his discovery of sedimentation equilibrium
1927	Arthur Holly Compton	1892–1962	for his discovery of the effect named after him
	Charles Thomson Rees Wilson	1869–1959	for his method of making the paths of electrically charged particles visible by condensation of vapor
1928	Owen Willans Richardson	1879–1959	for his work on the thermionic phenomenon and especially for the discovery of the law named after him
1929	Prince Louis-Victor de Broglie	1892–1987	for his discovery of the wave nature of electrons
1930	Sir Chandrasekhara Venkata Raman	1888–1970	for his work on the scattering of light and for the discovery of the effect named after him
1932	Werner Heisenberg	1901–1976	for the creation of quantum mechanics, the application of which has, among other things, led to the discovery of the allotropic forms of hydrogen
1933	Erwin Schrödinger	1887–1961	for the discovery of new productive forms of atomic theory
	Paul Adrien Maurice Dirac	1902–1984	
1935	James Chadwick	1891–1974	for his discovery of the neutron
1936	Victor Franz Hess	1883–1964	for the discovery of cosmic radiation
	Carl David Anderson	1905–1991	for his discovery of the positron
1937	Clinton Joseph Davisson	1881–1958	for their experimental discovery of the diffraction of electrons by crystals
	George Paget Thomson	1892–1975	
1938	Enrico Fermi	1901–1954	for his demonstrations of the existence of new radioactive elements produced by neutron irradiation, and for his related discovery of nuclear reactions brought about by slow neutrons
1939	Ernest Orlando Lawrence	1901–1958	for the invention and development of the cyclotron and for results obtained with it, especially for artificial radioactive elements
1943	Otto Stern	1888–1969	for his contribution to the development of the molecular ray method and his discovery of the magnetic moment of the proton
1944	Isidor Isaac Rabi	1898–1988	for his resonance method for recording the magnetic properties of atomic nuclei
1945	Wolfgang Pauli	1900–1958	for the discovery of the Exclusion Principle (Pauli Principle)
1946	Percy Williams Bridgman	1882–1961	for the invention of an apparatus to produce extremely high pressures, and for the discoveries he made therewith in the field of high-pressure physics
1947	Sir Edward Victor Appleton	1892–1965	for his investigations of the physics of the upper atmosphere, especially for the discovery of the so-called Appleton layer
1948	Patrick Maynard Stuart Blackett	1897–1974	for his development of the Wilson cloud chamber method, and his discoveries therewith in nuclear physics and cosmic radiation
1949	Hideki Yukawa	1907–1981	for his prediction of the existence of mesons on the basis of theoretical work on nuclear forces
1950	Cecil Frank Powell	1903–1969	for his development of the photographic method of studying nuclear processes and his discoveries regarding mesons made with this method
1951	Sir John Douglas Cockcroft	1897–1967	for their pioneer work on the transmutation of atomic nuclei by artificially accelerated atomic particles
	Ernest Thomas Sinton Walton	1903–1995	
1952	Felix Bloch	1905–1983	for their development of new methods for nuclear magnetic precision methods and discoveries of connection therewith
	Edward Mills Purcell	1912–1997	
1953	Frits Zernike	1888–1966	for his demonstration of the phase-contrast method, especially for his invention of the phase-contrast microscope
1954	Max Born	1882–1970	for his fundamental research in quantum mechanics, especially for his statistical interpretation of the wave function
	Walther Bothe	1891–1957	for the coincidence method and his discoveries made therewith
1955	Willis Eugene Lamb	1913–	for his discoveries concerning the fine structure of the hydrogen spectrum
	Polykarp Kusch	1911–1993	for his precision determination of the magnetic moment of the electron
1956	William Shockley	1910–1989	for their research on semiconductors and their discovery of the transistor effect
	John Bardeen	1908–1991	
	Walter Houser Brattain	1902–1987	
1957	Chen Ning Yang	1922–	for their penetrating investigation of the parity laws, which has led to important discoveries regarding the elementary particles
	Tsung Dao Lee	1926–	
1958	Pavel Aleksejecič Ĉerenkov	1904–1990	for the discovery and the interpretation of the Cerenkov effect
	Il' ja Michajlovič Frank	1908–1990	
	Igor Yevgenyevich Tamm	1895–1971	

1959	Emilio Gino Segrè	1905–1989	for their discovery of the antiproton
	Owen Chamberlain	1920–	
1960	Donald Arthur Glaser	1926–	for the invention of the bubble chamber
1961	Robert Hofstadter	1915–1990	for his pioneering studies of electron scattering in atomic nuclei and for his thereby achieved discoveries concerning the structure of the nucleons
	Rudolf Ludwig Mössbauer	1929–	for his researches concerning the resonance absorption of γ-rays and his discovery in this connection of the effect that bears his name
1962	Lev Davidoviĉ Landau	1908–1968	for his pioneering theories of condensed matter, especially liquid helium
1963	Eugene P. Wigner	1902–1995	for his contribution to the theory of the atomic nucleus and the elementary particles, particularly through the discovery and application of fundamental symmetry principles
	Maria Goeppert Mayer	1906–1972	for their discoveries concerning nuclear shell structure
	J. Hans D. Jensen	1907–1973	
1964	Charles H. Townes	1915–	for fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser–laser principle
	Nikolai G. Basov	1922–	
	Alexander M. Prochorov	1916–	
1965	Sin-itiro Tomonaga	1906–1979	for their fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles
	Julian Schwinger	1918–1994	
	Richard P. Feynman	1918–1988	
1966	Alfred Kastler	1902–1984	for the discovery and development of optical methods for studying Hertzian resonance in atoms
1967	Hans Albrecht Bethe	1906–	for his contributions to the theory of nuclear reactions, especially his discoveries concerning the energy production in stars
1968	Luis W. Alvarez	1911–1988	for his decisive contribution to elementary particle physics, in particular the discovery of a large number of resonance states, made possible through his development of the technique of using hydrogen bubble chamber and data analysis
1969	Murray Gell-Mann	1929–	for his contribution and discoveries concerning the classification of elementary particles and their interactions
1970	Hannes Alfvén	1908–1995	for fundamental work and discoveries in magneto-hydrodynamics with fruitful applications in different parts of plasma physics
	Louis Néel	1904–	for fundamental work and discoveries concerning antiferromagnetism and ferrimagnetism, which have led to important applications in solid state physics
1971	Dennis Gabor	1900–1979	for his discovery of the principles of holography
1972	John Bardeen	1908–1991	for their development of a theory of superconductivity
	Leon N. Cooper	1930–	
	J. Robert Schrieffer	1931–	
1973	Leo Esaki	1925–	for his discovery of tunneling in semiconductors
	Ivar Giaever	1929–	for his discovery of tunneling in superconductors
	Brian D. Josephson	1940–	for his theoretical prediction of the properties of a super-current through a tunnel barrier
1974	Antony Hewish	1924–	for the discovery of pulsars
	Sir Martin Ryle	1918–1984	for his pioneering work in radioastronomy
1975	Aage Bohr	1922–	for the discovery of the connection between collective motion and particle motion and the development of the theory of the structure of the atomic nucleus based on this connection
	Ben Mottelson	1926–	
	James Rainwater	1917–1986	
1976	Burton Richter	1931–	for their (independent) discovery of an important fundamental particle
	Samuel Chao Chung Ting	1936–	
1977	Philip Warren Anderson	1923–	for their fundamental theoretical investigations of the electronic structure of magnetic and disordered systems
	Nevill Francis Mott	1905–1996	
	John Hasbrouch Van Vleck	1899–1980	
1978	Peter L. Kapitza	1894–1984	for his basic inventions and discoveries in low-temperature physics
	Arno A. Penzias	1926–	for their discovery of cosmic microwave background radiation
	Robert Woodrow Wilson	1936–	
1979	Sheldon Lee Glashow	1932–	for their unified model of the action of the weak and electromagnetic forces for their prediction of the existence of neutral currents
	Abdus Salam	1926–1996	
	Steven Weinberg	1933–	
1980	James W. Cronin	1931–	for the discovery of violations of fundamental symmetry principles in the decay of neutral K mesons
	Val L. Fitch	1923–	

1981	Nicolaas Bloembergen	1920–	for their contribution of the development of laser spectroscopy
	Arthur Leonard Schawlow	1921–1999	
	Kai M. Siegbahn	1918–1999	for his contribution of high-resolution electron spectroscopy
1982	Kenneth Geddes Wilson	1936–	for his method of analyzing the critical phenomena inherent in the changes of matter under the influence of pressure and temperature
1983	Subrehmanyan Chandrasekhar	1910–1995	for his theoretical studies of the structure and evolution of stars
	William A. Fowler	1911–1995	for his studies of the formation of the chemical elements in the universe
1984	Carlo Rubbia	1934–	for their decisive contributions to the large project, which led to the discovery of the field particles W and Z, communicators of the weak interaction
	Simon van der Meer	1925–	
1985	Klaus von Klitzing	1943–	for his discovery of the quantized Hall resistance
1986	Ernst Ruska	1906–1988	for his invention of the electron microscope
	Gerd Binnig	1947–	for their invention of the scanning-tunneling electron microscope
	Heinrich Rohrer	1933–	
1987	Karl Alex Müller	1927–	for this discovery of a new class of superconductors
	J. Georg Bednorz	1950–	
1988	Leon M. Lederman	1922–	for experiments with neutrino beams and the discovery of the muon neutrino
	Melvin Schwartz	1932–	
	Jack Steinberger	1921–	
1989	Hans G. Dehmelt	1922–	for their development of technique for trapping individual atoms
	Wolfgang Paul	1913–1993	
	Norman F. Ramsey	1915–	for his discoveries in atomic resonance spectroscopy, which led to hydrogen masers and atomic clocks
1990	Richard E. Taylor	1929–	for their experiments on the scattering of electrons from nuclei, which revealed the presence of quarks inside nucleons
	Jerome I. Friedman	1930–	
	Henry W. Kendall	1926–1999	
1991	Pierre-Gilles de Gennes	1932–	for discoveries about the ordering of molecules in substances such as liquid crystals, superconductors, and polymers
1992	George Charpak	1924–	for his invention of fast electronic detectors for high-energy particles
1993	Joseph H. Taylor	1941–	for the discovery and interpretation of the first binary pulsar
	Russell A. Hulse	1950–	
1994	Bertram N. Brockhouse	1918–	for the development of neutron-scattering techniques
	Clifford G. Shull	1915–	
1995	Martin L. Perl	1927–	for the discovery of the tau lepton
	Frederick Reines	1918–1998	for the detection of the neutrino
1996	David M. Lee	1931–	for their discovery of superfluidity in ^{3}He
	Douglas M. Osheroff	1945–	
	Robert C. Richardson	1937–	
1997	Steven Chu	1948–	for the development of methods to cool and trap atoms with laser light
	Claude Cohen-Tannoudji	1933–	
	William D. Phillips	1948–	
1998	Robert B. Laughlin	1950–	for their discovery of a new form of quantum fluid with fractionally charged excitations
	Horst L. Stormer	1949–	
	Daniel C. Tsui	1939–	
1999	Gerardus 't Hooft	1946–	for elucidating the quantum structure of electroweak interactions in physics
	Martinus J. G. Veltman	1931–	
2000	Zhores I. Alferov	1930–	for developing semiconductor heterostructures used in high-speed and opto-electronics
	Herbert Kroemer	1928–	
	Jack S. Kilby	1923–	for his part in the invention of the integrated circuit

ANSWERS TO ODD-NUMBERED EXERCISES AND PROBLEMS

CHAPTER 25

Exercises

1. 0.50 C. **3.** 2.74 N. **5.** *(a)* 1.77 N. *(b)* 3.07 N. **7.** $q_1 = -4q_2$ **9.** 24.5 N, along the angle bisector. **11.** At the point 82.3 cm from the positive charge and 144 cm from the negative charge. **15.** *(a)* $\vec{\mathbf{F}} = F_z\hat{\mathbf{k}}$. *(b)* $(2q_0q/4\pi\epsilon_0R^2)[1 - [z^2/(z^2 + R^2)]^{1/2}](z/|z|)\hat{\mathbf{k}}$. **17.** Along the y axis a distance of $(y^4 + y^2L^2/4)^{1/4}$ *(a)* to the right of q_1 or *(b)* to the left of q_0. **19.** $F_z = -(q_0\lambda/2\pi\epsilon_0)[1/y - 1/(y^2 + L^2/4)^{1/2}]$. **21.** *(a)* Boron. *(b)* Nitrogen. *(b)* Carbon. **23.** 2.89 nN. **25.** 3.8 N. **27.** 5.08 m below the electron. **29.** 13.4 MC. **31.** *(a)* 57.1 TC; no. *(b)* 598 metric tons.

Problems

1. 1.00 μC and 3.00 μC, but of opposite sign. **3.** *(a)* A charge $-4q/9$ must be located on the line segment joining the two positive charges, a distance $L/3$ from the $+q$ charge. **5.** *(b)* 2.96 cm. **7.** $q = Q/2$. **9.** $(\pi^3 m\epsilon_0 d^3/2qQ)^{1/2}$. **11.** $a/\sqrt{2}$.

CHAPTER 26

Exercises

1. 10.5 mN/C, westward. **3.** 203 nN/C, up. **5.** 144 pC. **7.** 19.5 kN/C. **9.** 9:30. **13.** $R/\sqrt{3}$. **15.** *(a)* 104 nC. *(b)* 1.31×10^{17}. *(c)* 4.96×10^{-6}. **17.** *(a)* 6.50 cm. *(b)* 4.80 μC. **23.** To the right. **27.** *(a)* 6.53 cm. *(b)* 26.9 ns. *(c)* 0.121. **29.** *(a)* 585 kN/C, toward the negative charge. *(b)* 93.6 fN, toward the positive charge. **31.** $5e$. **33.** 1.64×10^{-19} C ($\approx$ 2.5% high). **35.** 1.2 mm. **37.** *(a)* Zero. *(b)* 8.50×10^{-22} N · m. *(c)* Zero.

Problems

1. *(b)* Parallel to $\vec{\mathbf{p}}$. **3.** *(a)* $qz/4\pi\epsilon_0(R^2 + z^2)^{3/2}$. *(b)* $(q_1 - q_2)R/2\pi^2\epsilon_0(R^2 + z^2)^{3/2}$. **5.** $+q$ at $z = +2a$ and $z = -2a$; $-4q$ at $z = +a$ and $z = -a$; $+6q$ at $z = 0$. **11.** The upper plate; 4.06 cm. **13.** $2pE\cos\theta_0$. **15.** *(a)* $8q/\pi\epsilon_0a^3$.

CHAPTER 27

Exercises

1. -7.8 mN · m²/C. **3.** *(a)* $-\pi R^2E$. *(b)* πR^2E. **5.** 208 kN · m²/C. **7.** $q/6\epsilon_0$. **9.** 4.6 μC. **13.** *(a)* 322 nC. *(b)* 143 nC. **15.** *(a)* Zero. *(b)* σ/ϵ_0, to the left. *(c)* Zero. **17.** 5.09 μC/m³. **19.** -1.13 nC. **21.** *(a)* 2.19 MN/C, radially out. *(b)* 436 kN/C, radially in. **23.** 97.9 cm. **25.** *(b)* $\rho R^2/2\epsilon_0 r$. **27.** *(a)* 452 nC/m². *(b)* 51.1 kN/C. **29.** *(a)* 53 MN/C. *(b)* 60 N/C.

Problems

3. 5.11 nC/m². **5.** *(a)* $q/2\pi\epsilon_0Lr$, radially inward. *(b)* $-q$ on both inner and outer surfaces. *(c)* $q/2\pi\epsilon_0Lr$, radially outward. **7.** *(a)* $\lambda/2\pi\epsilon_0r$. *(b)* Zero. **9.** 270 eV. **11.** *(a)* $q/4\pi\epsilon_0r^2$. *(b)* $q/4\pi\epsilon_0r^2$. *(c)* Only within the shell itself. *(d)* Yes, charges are induced on the surfaces. *(e)* Yes. *(f)* No. *(g)* No. **13.** $q/2\pi a^2$. **19.** *(a)* $-Q$. *(b)* $-Q$. *(c)* $-(Q + q)$. *(d)* Yes.

CHAPTER 28

Exercises

1. *(a)* 484 keV. *(b)* Zero. **3.** *(a)* 27.2 fJ. *(b)* 3.02×10^{-31} kg, in error by a factor of about 3. **5.** *(a)* 3.0 kN. *(b)* 240 MeV. **7.** *(a)* 30 GJ. *(b)* 7.1 km/s. *(c)* 9.0×10^4 kg. **9.** 2.6 km/s. **11.** *(a)* 27.2 V. *(b)* -27.2 eV. *(c)* 13.6 eV. *(d)* 13.6 eV. **13.** *(a)* 24.4 kV/m. *(b)* 2.93 kV. **15.** *(a)* 132 MV/m. *(b)* 8.43 kV/m. **17.** *(a)* 32 MeV. **19.** *(a)* -3.85 kV. *(b)* -3.85 kV. **21.** -1.1 nC. **23.** 637 MV. **25.** *(a)* $qd/2\pi\epsilon_0a(a + d)$. **27.** *(a)* 2.76 MV. *(b)* 5.27 J. *(c)* 7.35 J. **29.** 186 pJ. **31.** 746 V/m. **33.** -2.3×10^{21} V/m. **35.** -39.2 V/m. **37.** *(a)* 2.46 V. *(b)* 2.46 V. *(c)* Zero. **43.** *(a)* -115 mV. *(b)* 18.1 nV/m. **45.** *(a)* 3.64 nC. *(b)* 12.5 nC/m². **47.** *(a)* Large sphere: 38.6 nC; small sphere: 18.6 nC. *(b)* 2.84 kV each. **49.** *(a)* 110 μC; 1.1 μC. *(b)* 8.75 μC/m²; 8.75 μC/m².

Problems

1. *(a)* 256 kV. *(b)* $0.745c$. **3.** $(qQ/8\pi\epsilon_0)(1/r_1 - 1/r_2)$. **5.** 2.17 d. **7.** *(a)* 562 μm. *(b)* 813 V. **9.** *(a)* -5.40 nm. *(b)* 9.00 nm. *(c)* No. **13.** *(a)* $(k/4\pi\epsilon_0)[(L^2 + y^2)^{1/2} - y]$. *(b)* $(k/4\pi\epsilon_0)[1 - y/(L^2 + y^2)^{1/2}]$. *(c)* $0.75L$. **15.** 2.0×10^{-8}. **17.** *(a)* Zero. *(b)* Zero. *(c)* Zero. *(d)* Zero. *(e)* No.

CHAPTER 29

Exercises

1. *(a)* 1.33 kC. *(b)* 8.31×10^{21}. **3.** *(a)* 9.41 A/m², north. **5.** 0.400 mm. **7.** 0.67 A, toward the negative terminal. **9.** *(a)* 654 nA/m². *(b)* 83.4 MA. **11.** 52.5 min. **13.** 0.59 Ω. **15.** *(a)* 1.5 kA. *(b)* 53 MA/m². *(c)* 110 nΩ · m; platinum. **19.** 3. **21.** *(a)* 6.00 mA. *(b)* 15.9 nV. *(c)* 21.2 nΩ. **23.** 1190 $(\Omega \cdot \text{m})^{-1}$. **25.** *(a)* Cu: 55.3 A/cm²; Al: 34.0 A/cm². *(b)* Cu: 1.01 kg; Al: 495 g. **27.** *(a)* 8.52 kΩ. *(b)* 4.51 μA. **29.** *(a)* 7.65×10^5 N/C. *(b)* 3.60×10^6 N/C. *(c)* 2.51×10^{-5} C/m². **31.** *(a)* $(3 \times 10^6$ V/m$)R$

Problems

1. 7.1 ms. **3.** *(a)* 380 μV. *(b)* Negative. *(c)* 4.3 min. **5.** *(a)* 95.0 μC. *(b)* 158 C°. **7.** *(a)* 250°C. **9.** 54 Ω. **11.** *(a)* Silver. *(b)* 60.8 nΩ. **13.** 0.036. **15.** $R = (\rho/4\pi)(1/a - 1/b)$.

CHAPTER 30

Exercises

1. 7.5 pC. **3.** 3.25 mC. **5.** 0.546 pF. **7.** *(a)* 84.5 pF. *(b)* 191 cm^2. **9.** 9090. **11.** 7.17 μF. **13.** *(a)* 2.4 μF. *(b)* $q_4 = q_6 = 480\ \mu\text{C}$. *(c)* $\Delta V_4 = 120\ \text{V}$; $\Delta V_6 = 80\ \text{V}$. **15.** *(a)* $d/3$. *(b)* $3d$. **17.** *(a)* 942 μC. *(b)* 91.4 V. **21.** 13.2 ¢. **23.** *(a)* 28.6 pF. *(b)* 17.9 nC. *(c)* 5.59 μJ. *(d)* 482 kV/m. **25.** 74.1 mJ/m^3. **29.** 3.89. **31.** The mica sheet. **33.** 86.3 nF. **35.** *(a)* 730 pF. *(b)* 28 kV. **37.** *(a)* $\epsilon_0 A/(d - b)$. *(b)* $d/(d - b)$. *(c)* $q^2 b/2A\epsilon_0$; pulled in. **39.** *(a)* 13.4 kV/m. *(b)* 6.16 nC. *(c)* 5.02 nC.

Problems

5. *(a)* 45.4 V. *(b)* 52.7 μC. *(c)* 146 μC. **7.** *(a)* 50 V. *(b)* Zero. **9.** *(a)* $q_1 = 9.0\ \mu\text{C}$; $q_2 = 16\ \mu\text{C}$; $q_3 = 9.0\ \mu\text{C}$; $q_4 = 16\ \mu\text{C}$. *(b)* $q_1 = 8.40\ \mu\text{C}$; $q_2 = 16.8\ \mu\text{C}$; $q_3 = 10.8\ \mu\text{C}$; $q_4 = 14.4\ \mu\text{C}$. **11.** *(a)* 2.0 J. **13.** *(a)* $e^2/32\pi^2\epsilon_0 r^4$. *(b)* $e^2/8\pi\epsilon_0 R$. *(c)* 1.40 fm. **17.** 1.63 kV. **21.** *(a)* $2\Delta V$. *(b)* $U_i = \epsilon_0 A(\Delta V)^2/2d$; $U_f = \epsilon_0 A(\Delta V)^2/d$. *(c)* $\epsilon_0 A(\Delta V)^2/2d$. **23.** *(a)* 85.6 pF. *(b)* 119 pF. *(c)* 10.3 nC; 10.3 nC. *(d)* 9.86 kV/m. *(e)* 2.05 kV/m. *(f)* 86.6 V. *(g)* 170 nJ.

CHAPTER 31

Exercises

1. 10.6 kJ. **3.** 13 h 38 min. **5.** -10 V. **7.** *(a)* 14 Ω. *(b)* 35 mW. **9.** *(a)* 44.2 V. *(b)* 21.4 V. *(c)* Left. **11.** $\mathcal{E}/7R$. **13.** *(a)* $i_1 = 668$ mA, down: $i_2 = 85.7$ mA, up; $i_3 = 582$ mA, up. *(b)* -3.60 V. **15.** *(a)* 3.4 A. *(b)* 0.29 V. And: *(a)* 0.59 A. *(b)* 1.7 V. **17.** 4.0 Ω; 12 Ω. **19.** 7.5 V. **21.** 262 Ω or 38.2 Ω. **23.** *(a)* 131 Ω. *(b)* $i_1 = 47.5$ mA; $i_2 = 21.2$ mA; $i_3 = 14.4$ mA; $i_4 = 11.9$ mA. **25.** *(a)* $R/2$. *(b)* $5R/8$. **29.** 18 kC. **31.** *(a)* 1.03 kW. *(b)* 34.5 ¢. **33.** *(a)* R_2. *(b)* R_1. **35.** *(a)* \$4.46. *(b)* 144 Ω. *(c)* 833 mA. **37.** *(a)* 2.88×10^{11}. *(b)* 24.0 μA. *(c)* 1.14 kW; 23.1 mW. **39.** *(a)* 6.1 m. *(b)* 13 m. **43.** 4.61. **45.** *(a)* 2.20 s. *(b)* 44 mV. **47.** 2.35 MΩ. **49.** *(a)* 955 nA. *(b)* 1.08 μW. *(c)* 2.74 μW. *(d)* 3.82 μW.

Problems

1. The cable. **3.** *(a)* 1.5 kΩ. *(b)* 400 mV. *(c)* 0.26%. **5.** *(a)* $\rho_A = 16.3$ n$\Omega\cdot$m; $\rho_B = 7.48$ n$\Omega\cdot$m. *(b)* $j_A = j_B = 62.3$ kA/cm^2. *(c)* $E_A = 10.2$ V/m; $E_B = 4.66$ V/m. *(d)* $\Delta V_A = 435$ V; $\Delta V_B = 195$ V. **7.** *(a)* $3R/4$. *(b)* $5R/6$. **9.** *(a)* In parallel. *(b)* 72.0 Ω; 144 Ω. **11.** 0.45 A. **15.** 27.4 cm/s. **17.** *(a)* $1.37L$. *(b)* $0.730A$. **19.** $RC \ln 2$.

CHAPTER 32

Exercises

1. 1: +; 2: −; 3: 0; 4: −. **3.** *(a)* 3.4 km/s. **5.** 8.2×10^9. **9.** *(a)* 0.34 mm. *(b)* 2.6 keV. **11.** *(a)* 1.11×10^7 m/s. *(b)* 0.316 mm. **13.** *(a)* 2600 km/s. *(b)* 110 ns. *(c)* 140 keV. *(d)* 70 kV. **15.** *(a)* K_p. *(b)* $K_p/2$. **17.** *(a)* $\sqrt{2}r_p$. *(b)* r_p. **19.** *(a)* $0.999928c$. **21.** An alpha particle. **23.** ≈ 240 m. **25.** 37 cm/s. **29.** 467 mA; left to right. **31.** *(a)* 330 MA. *(b)* 1.1×10^{17} W. **33.** $(-0.414\ \text{N})\hat{\mathbf{k}}$ **35.** *(a)* Zero; 0.138 N; 0.138 N. **37.** *(a)* 20 min. *(b)* 0.059 N$\cdot$m.

Problems

1. $(0.75\ \text{T})\hat{\mathbf{k}}$ **3.** *(a)* To the east. *(b)* 6.27×10^{14} m/s^2. *(c)* 2.98 mm. **7.** *(a)* $B\Delta x(qm/2\Delta V)^{1/2}$. *(b)* 7.91 mm. **9.** *(a)* $-q$. *(b)* $\pi m/qB$. **11.** *(a)* 78.6 ns. *(b)* 9.16 cm. *(c)* 3.20 cm. **15.** 4.2 C. **19.** 1.63 A.

CHAPTER 33

Exercises

1. *(a)* 3×10^8 m/s. **3.** 7.7 mT. **5.** 12 nT. **7.** *(a)* 0.324 fN, parallel to current. *(b)* 0.324 fN, radially outward. *(c)* Zero. **9.** 30.0 A, antiparallel. **11.** *(a)* 2.43 A$\cdot$m^2. *(b)* 46 cm. **13.** $(\mu_0 i\theta/4\pi)(1/b - 1/a)$, out of page. **15.** $(\mu_0 i/2\pi w)\ln(1 + w/d)$, up. **19.** *(a)* 68 μT. *(b)* 9.3×10^{-7} N$\cdot$m. **21.** 606 μN, toward center of square. **23.** *(b)* 2.3 km/s. **27.** 109 m. **31.** *(a)* -2.5 μT$\cdot$m. *(b)* Zero. **33.** *(a)* $\mu_0 ir/2\pi c^2$. *(b)* $\mu_0 i/2\pi r$. *(c)* $(\mu_0 i/2\pi r)[(a^2 - r^2)/(a^2 - b^2)]$. *(d)* Zero. **35.** $3i_0/8$, into the page. **37.** *(a)* Negative. *(b)* 9.7 cm.

Problems

1. $8N\mu_0 i/5\sqrt{5}R$. **5.** *(c)* $\frac{1}{2}nia^2 \sin(2\pi/n)$. **7.** *(a)* $(2\mu_0 i/3\pi L)(2\sqrt{2} + \sqrt{10})$. *(b)* Greater. **9.** 0.272 A. **13.** $\mu_0 ir^2/2\pi a^3$.

CHAPTER 34

Exercises

1. 57 μWb. **3.** *(a)* 31 mV. *(b)* Right to left. **5.** *(a)* 1.12 mΩ. *(b)* 1.27 T/s. **7.** *(b)* 58 mA. **9.** 4.97 μW. **11.** *(b)* No. **13.** 600 nV. **15.** Zero. **17.** $iLBt/m$, away from G. **21.** 25 μC. **23.** *(a)* $(\mu_0 ia/2\pi)\ln(1 + b/D)$. *(b)* $\mu_0 iabv/2\pi RD(D + b)$. **25.** 6.3 rev/s. **27.** 5.5 kV. **29.** *a*: -1.20 mV; *b*: -2.79 mV; *c*: 1.59 mV. **31.** *(a)* 4.53×10^7 m/s^2, to the right. *(b)* Zero. *(c)* 4.53×10^7 m/s^2, to the left. **33.** *(a)* 0.15°.

Problems

3. *(a)* 28.2 μV. *(b)* From *c* to *b*. **5.** 80 μV, clockwise. **7.** 0.455 V. **11.** $(Bar)^2\omega\sigma t$. **15.** *(a)* 34 V/m. *(b)* 6.0×10^{12} m/s^2.

CHAPTER 35

Exercises

1. 2.1×10^9 A. **3.** *(a)* $(-2.86\ \text{A}\cdot\text{m}^2)\hat{\mathbf{k}}$. *(b)* $(1.10\ \text{A}\cdot\text{m}^2)\hat{\mathbf{k}}$. **5.** *(b)* ia^2. **7.** $\frac{1}{2}\pi i(a^2 + b^2)$. **9.** *(a)* 514 GV/m. *(b)* 19.0 mT. **13.** 24 mJ/T. **15.** 0.58 K. **17.** *(a)* 150 T. *(b)* 450 T. **19.** Yes. **21.** *(a)* 3.0 μT. *(b)* 9.0×10^{-29} J. **23.** *(a)* 180 km. *(b)* 2.3×10^{-5}. **25.** 1660 km. **27.** 61 μT; 84°. **29.** $+3$ Wb. **31.** $(\mu_0 iL/\pi)\ln 3$.

Problems

3. *(b)* K_i/B, opposite to the field. *(b)* 312 A/m.

CHAPTER 36

Exercises

1. 100 nWb. **3.** 0.261 mH/m. **5.** *(a)* 0.60 mH. *(b)* 120. **7.** 7.87 H. **11.** 29.8 Ω. **13.** *(a)* 4.78 mH. *(b)* 2.42 ms.

13. (*a*) 4.78 mH. (*b*) 2.42 ms. **15.** 42 V + (20 V/s)t.
17. 12 A/s. **19.** (*a*) $i_1 = i_2 = 3.33$ A. (*b*) $i_1 = 4.55$ A;
$i_2 = 2.73$ A. (*c*) $i_1 = 0$; $i_2 = 1.82$ A. (*d*) $i_1 = i_2 = 0$.
21. I. (*a*) 2.0 A. (*b*) Zero. (*c*) 2.0 A. (*d*) Zero. (*e*) 10 V.
(*f*) 2.0 A/s. II. (*a*) 2.0 A. (*b*) 1.0 A. (*c*) 3.0 A. (*d*) 10 V.
(*e*) Zero. (*f*) Zero. **23.** (*a*) 13.2 H. (*b*) 0.124 A.
25. 63.2 MJ/m^3. **27.** 150 MV/m. **29.** (*a*) 117 H. (*b*) 0.225 mJ.
31. 12×10^{15} J. **33.** 0.123 A. **35.** 0.038 mH.
39. (*a*) 89.3 rad/s. (*b*) 70.3 ms. (*c*) 24.1 μF. **41.** (*a*) 6.08 μs.
(*b*) 164 kHz. (*c*) 3.04 μs. **43.** (*a*) No. (*b*) 6.1 kHz. (*c*) 16 nF.
45. (*a*) 5800 rad/s. (*b*) 1.1 ms. **47.** (*a*) $q_m/\sqrt{3}$.
(*b*) $t/T = 0.152$. **49.** (*a*) 6.0 : 1. (*b*) 36 pF; 220 μH.
51. (*a*) 180 μC. (*b*) $T/8$. (*c*) 67 W. **53.** $(L/R) \ln 2$.
55. 2.96 Ω.

Problems

9. (*a*) $\mu_0 i^2 N^2 / 8\pi^2 r^2$. (*b*) $(\mu_0 N^2 h i^2 / 4\pi) \ln (b/a)$. **13.** (*a*) Zero.
(*b*) $2i$.

CHAPTER 37

Exercises

1. 377 rad/s. **3.** (*a*) 3750 rad/s. (*b*) 23.4 Ω.
5. (*a*) 39.1 mA. (*b*) Zero. (*c*) 32.6 mA. (*d*) Supplying energy.
11. 1.0 kV ($> \mathcal{E}_m$). **13.** (*a*) 36.0 V. (*b*) 27.4 V. (*c*) 17.0 V.
(*d*) 8.4 V. **15.** (*a*) 39.1 Ω. (*b*) 21.7 Ω. (*c*) Capacitive.
17. 177 Ω. **19.** (*a*) 1.82 W. (*b*) 3.13 W. **21.** 100 V.
25. (*a*) 2.49 A. (*b*) 37.4, 153, 218, 65.0, 75.0 V.
(*c*) $P_C = P_L = 0$; $P_R = 93.0$ W. **27.** 1.8 kV.
29. Step-up: 5; 4; 1.25. Step-down: 0.8; 0.25; 0.2. **31.** 40 V.

Problems

1. (*a*) 6.73 m/s. (*b*) 11.2 ms. (*c*) Inductor. (*d*) 144 mH.
3. (*a*) 45°. (*b*) 76.0 Ω. **7.** (*a*) 229.0 rad/s. (*b*) 6.11 A.
(*c*) 233.5, 224.5 rad/s. (*d*) 0.039. **9.** (*a*) 41.2 W. (*b*) -16.9 W.
(*c*) 43.7 W. (*d*) 14.3 W. **11.** (*a*) 76.4 mH. (*b*) 17.8 Ω.

CHAPTER 38

Exercises

1. $r = 2.5$ cm, 10 cm. **3.** Change the potential across the plate at the rate 1.0 kV/s. **5.** (*a*) 1.84 A. (*b*) 140 GV/m · s.
(*c*) 460 mA. (*d*) 578 nT · m. **7.** (*a*) 0.84 A. (*b*) Zero.
(*c*) 1.3 A. **9.** 2.27 pT. **15.** 5.0×10^{-21} H. **17.** 1.07 pT.
21. 100 kJ. **23.** 4.62×10^{-29} W/m^2. **25.** 78 cm.
27. (*a*) 883 m. (*b*) No. **29.** (*a*) 6.53 nT. (*b*) 5.10 mW/m^2.
(*c*) 8.04 W. **31.** (*a*) 76.8 mV/m. (*b*) 256 pT. (*c*) 12.6 kW.
33. 0.043 kg · m/s. **35.** 7.7 MPa. **37.** (*a*) 586 MN.
(*b*) 1.66×10^{-14}. **39.** $I(2 - f)/c$. **43.** 490 nm.

Problems

3. (*a*) 7.63 μA. (*b*) 862 kV · m/s. (*c*) 3.48 mm. (*d*) 5.07 pT.
5. (*a*) $\pm Eba^2/\mu_0$ for faces parallel to the xy plane; zero through each of the other four faces. (*b*) Zero.
9. (*a*) $\omega/k = c$; $E_m = cB_m$. (*b*) $S = (E_m{}^2/4\mu_0 c) \sin 2\omega t \sin 2kx$.
11. (*a*) $E = \mathcal{E}/r \ln (b/a)$; $B = \mu_0 \mathcal{E}/2\pi R r$.
(*b*) $S = \mathcal{E}^2/2\pi R r^2 \ln (b/a)$. **13.** (*a*) 94.3 MHz. (*b*) $+z$; 960 nT.
(*c*) 1.98 m^{-1}; 593 Mrad/s. (*d*) 110 W/m^2. (*e*) 678 nN; 367 nPa.
15. (*a*) 3.60 GW/m^2. (*b*) 12.0 Pa. (*c*) 16.7 pN.
(*d*) 2.78 km/s^2.

CHAPTER 39

Exercises

3. (*a*) 4.5×10^{24} Hz. (*b*) 10,000 km. **5.** (*a*) 515 nm; 610 nm, approximately. (*b*) 555 nm; 541 THz; 1.85 fs. **7.** (*a*) 8.68 y.
(*b*) 4.4 My. **9.** 67 ps. **13.** (*a*) 38.0°. (*b*) 52.9°. **15.** 1.56.
17. 1.95×10^8 m/s. **19.** 1.25. **23.** 74 m. **27.** 43 mm.
29. (*a*) 405 nm. (*b*) 2.37 μm. (*c*) 42.5°. **33.** (*a*) 72.07°.
(*b*) From A to B. **35.** 187 cm. **37.** (*a*) Yes. (*b*) No. (*c*) 43°.
41. Yellow-orange. **43.** (*b*) 0.80c. **45.** ± 0.0036 nm.
47. (*a*) 1.66×10^{-5}. (*b*) 0.83×10^{-5}. **51.** 78.9°.

Problems

3. 22°. **5.** (*b*) 0.60 mm. **7.** 750 m. **9.** (*a*) $n_{\text{liquid}} < n_{\text{glass}}$.
11. (*b*) 0.170. **13.** (*b*) 60.2 μs. **15.** 59.4 m/s.

CHAPTER 40

Exercises

1. (*a*) $2v$. (*b*) v. **7.** 390 cm beneath the mirror surface.
9. 1.7 mm^2. **11.** Six. **13.** 11.0 cm.
15. (*a*) +, +40, −20, +2, no, yes. (*b*) Plane, ∞, ∞, −10, yes.
(*c*) Concave, +40, +60, −2, yes, no. (*d*) Concave, +20, +40, +30, yes, no. (*e*) Convex, −20, +20, +0.50, no, yes.
(*f*) Convex, −, −40, −18, +180, no, yes. (*g*) −20, −, −, +5.0, +0.80, no, yes. (*h*) Concave, +8.0, +16, +12, −, yes.
19. (*b*) 2.0. (*c*) None. **21.** 12 cm to the left of the lens.
23. 2.5 mm. **25.** (*a*) 40 cm. (*b*) 80 cm. (*c*) 240 cm.
(*d*) −40 cm. (*e*) −80 cm. (*f*) −240 cm. **29.** 22 cm.
31. 30 cm to the left of the diverging lens; virtual; upright; $m = 1.0$. **33.** (*a*) 73.6 cm on the side of the lens away from the mirror. (*b*) Real. (*c*) Upright. (*d*) 0.289. **35.** 2.0 mm.
37. (*a*) 2.34 cm. (*b*) Smaller. **39.** 103.

Problems

1. $I_{\text{new}} = (10/9) I_{\text{old}}$. **9.** (*b*) Separate the lenses by $f_2 - |f_1|$.
(*c*) $(f_1/f_2)^2$. **13.** (*a*) 5.3 cm. (*b*) 3.0 mm. **15.** 25 ms.

CHAPTER 41

Exercises

1. (*a*) 0.22 rad. (*b*) 12°. **3.** 2.3 mm. **5.** 650 nm.
7. 0.103 mm. **9.** 600 nm. **13.** 3.2×10^{-4}. **15.** 0°.
19. (*a*) 1.21, 3.22, 8.13 m. **21.** 124 nm. **23.** (*a*) 552 nm.
(*b*) 442 nm. **25.** 215 nm. **27.** 840 nm. **29.** 141.
31. 1.89 μm. **33.** (*a*) 34. (*b*) 45. **35.** 1.00 m. **37.** (*a*) 88%.
(*b*) 95%. **39.** 588 nm.

Problems

1. (*a*) 0.253 mm. (*b*) Maxima replace minima and minima replace maxima. **7.** 643 nm. **9.** 2.4 μm.

CHAPTER 42

Exercises

1. 690 nm. **3.** (*a*) 0.430°. (*b*) 118 μm. **5.** (*a*) $\lambda_a = 2\lambda_b$.
(*b*) Minima coincide when $m_b = 2m_a$. **7.** 173 μm. **9.** 1.49 mm.
11. (*a*) 0.186°. (*b*) 0.478 rad. (*c*) 0.926. **13.** 5.07°.
15. (*a*) 0.137 mrad. (*b*) 10.4 km. **17.** 51.8 m.
19. 1400 km. **21.** 15 m. **23.** (*a*) 6.8°.
(*b*) There is no minimum. **25.** $\lambda D/d$. **27.** (*a*) Three.
29. (*a*) Nine. (*b*) 0.255.

Problems

1. 24.4 mm. **3.** *(b)* 0, 4.493, 7.725 rad, etc.
(c) − 0.50, 0.93, 1.96, etc. **5.** *(b)* 70 μm.
(c) Three times the lunar diameter. **7.** *(a)* 5.0 μm. *(b)* 20 μm.

CHAPTER 43

Exercises

1. *(a)* 3.50 μm. *(b)* 9.69°; 19.7°; 30.3°; 42.3°; 57.3°.
3. 523 nm. **5.** *(a)* 6.0 μm. *(b)* 1.5 μm.
(c) $m = 0, 1, 2, 3, 5, 6, 7, 9$. **7.** 400 nm $< \lambda <$ 635 nm.
9. Three. **15.** 470 nm $< \lambda <$ 560 nm. **17.** 491.
19. 3650. **21.** 9.98 μm. *(b)* 3.27 mm.
23. *(a)* 0.032°/nm; 0.077°/nm; 0.25°/nm.
(b) 40,000; 80,000; 120,000. **25.** 2.68°. **27.** 26 pm; 39 pm.
29. 49.8 pm. **33.** 0.206 nm.

Problems

3. *(b)* Half-way between principal maxima. *(c)* $I_m/9$.
7. *(a)* $a_0/\sqrt{2}$; $a_0/\sqrt{5}$; $a_0/\sqrt{10}$; $a_0/\sqrt{13}$; $a_0/\sqrt{17}$.

CHAPTER 44

Exercises

1. *(a)* $-y$. *(b)* $E_x = 0$; $E_y = 0$; $E_z = -cB \sin(ky + \omega t)$.
(c) Linearly polarized; z direction. **3.** *(a)* 2.14 V/m.
(b) 20.3×10^{-12} Pa. **5.** 1/8. **7.** 27/128. **9.** 15.8 W/m^2.
11. *(a)* 53.1°. *(b)* Yes, slightly. **13.** 55°31′ to 55°46′.
15. *(a)* Turns plane of polarization by 90°.
(b) Reverses handedness of circular polarization.
(c) Light remains unpolarized.

Problems

1. *(a)* 0.16. *(b)* 0.84. **3.** $\frac{1}{2}[\cos^2(\theta/N)]^N \rightarrow 0.5$ as $N \rightarrow \infty$.
5. *(a)* Polarizing sheet. *(b)* 45°.

CHAPTER 45

Exercises

1. *(b)* 2.11 eV. **3.** 1.17 eV. **5.** Ultraviolet.
7. *(a)* The infrared bulb. *(b)* 1.97×10^{20}. **9.** 91 K.
11. *(a)* 1.06 mm; microwave. *(b)* 9.4 μm; infrared.
(c) 1.6 μm; infrared. *(d)* 500 nm; visible. *(e)* 0.29 nm; x ray.
(f) 2.9×10^{-41} m; hard gamma ray. **13.** 580 mW.
15. *(a)* 7650 K. *(b)* 17,200 K. **17.** Cesium, lithium, barium.
19. *(a)* No. *(b)* 544 nm; green. **21.** 172 nm. **23.** *(a)* 1.17 V.
(b) 641 km/s. **27.** *(a)* 29.8 keV. *(b)* 7.19×10^{18} Hz.
(c) 1.59×10^{-23} kg · m/s = 29.8 keV/c. **29.** 2.95 cm/s.
31. *(a)* 2.87 pm. *(b)* 5.89 pm. **35.** 2.64 fm. **37.** *(a)* 4.86 pm.
(b) − 42.1 keV. *(c)* 42.1 keV. **39.** 42.6°.
41. *(a)* 588.9944 nm. *(b)* 299.97 m/s. *(c)* 299.94 m/s.

Problems

1. 1.44 W. **3.** 1300 K. **5.** *(b)* 6°C. **7.** *(a)* 2.97×10^{20} s^{-1}.
(b) 48,600 km. *(c)* 281 m. *(d)* 5.91×10^{18} (m$^2 \cdot$ s)$^{-1}$;
1.97×10^{10} m^{-3}. **9.** *(a)* 3.10 keV. *(b)* 14.4 keV.
11. *(b)* 1.12 keV.

CHAPTER 46

Exercises

1. *(a)* 1.7×10^{-35} m. **3.** *(a)* 38.8 pm. *(b)* 1.24 nm.
(c) 907 fm. **5.** *(a)* 3510 km/s. *(b)* 64.4 kV. **7.** *(a)* 5.3 fm.
9. 0.025 fm. **11.** A neutron. **13.** *(a)* 7.77 pm. *(b)* 7.68 pm.
15. 5.5°. **17.** *(a)* The beams are not present. *(b)* 47°.
19. 0.69 Hz. **21.** 76 μeV. **23.** 8.8×10^{-24} kg · m/s.
25. $\lambda/2\pi$. **27.** *(a)* 1.1×10^{-20} kg · m/s. *(b)* 19 MeV.
29. *(a)* 3.0×10^{-5}. *(b)* 2.5×10^{-7}.

Problems

1. 5.07 eV.

CHAPTER 47

Exercises

1. *(a)* 1900 MeV. *(b)* 1.0 MeV; no. **3.** 88.3 eV.
5. *(a)* 6.2×10^{-41} J. *(b)* 1.0×10^{-20}. *(c)* 3.0×10^{-18} K.
7. *(a)* 8.74 keV. *(b)* 1.01×10^{-22} kg · m/s. *(c)* 98.5 pm.
9. *(a)* $x = Nl/2n$, $N = 1, 3, 5, \ldots, (n-1)$.
(b) $x = Nl/n$, $N = 0, 1, 2, \ldots, n$. **13.** 656.3 nm; 486.1 nm;
434.1 nm; 410.2 nm; 397.0 nm. **15.** 3.40 eV.
17. *(a)* $n = 5 \rightarrow 3$. *(b)* Paschen. **19.** 66 neV; $E_2 = 3.4$ eV.
27. *(b)* 1013 nm; 364.6 nm. *(c)* Ultraviolet, visible, infrared.
29. *(b)* n^2. *(c)* n. *(d)* $1/n$. *(e)* $1/n^3$. *(f)* $1/n$. *(g)* $1/n^4$.
(h) $1/n^4$. *(i)* $1/n^2$. *(j)* $1/n^2$. *(k)* $1/n^2$.
31. *(a)* 30.4 nm; 292 nm; 1055 nm.
(b) 823 THz; 366 THz; 206 THz. **33.** *(a)* 2150 nm^{-3}; zero.
(b) 291 nm^{-3}; 10.2 nm^{-1} **35.** 1.85.
37. *(a)* $3, 2, 1, 0, -1, -2, -3\,\hbar$. *(b)* $-3, -2, -1, 0, 1, 2, 3$
μ_B. *(c)* 30.0°; 54.7°; 73.2°; 90.0°; 106.8°; 125.3°; 150.0°.
(d) $\sqrt{12}\,\hbar$. *(e)* $\sqrt{12}\,\mu_B$. **41.** *(a)* $0.764a_0$; $5.236a_0$.
(b) 0.981 nm^{-1}; 3.61 nm^{-1}. **43.** $n = 4$; $l = 3$;
$m_l = 3, 2, 1, 0, -1, -2, -3$; $m_s = \pm\frac{1}{2}$. **45.** $n \geq 5$;
$l = 4$; $m_s = \pm\frac{1}{2}$. **47.** $1, 0, 0, \frac{1}{2}$; $1, 0, 0, -\frac{1}{2}$.
49. All the statements are true.

Problems

1. 18.1, 36.2, 54.3, 66.3, 72.4 μeV. **3.** *(b)* 0.0006. *(c)* 0.0003.
5. $\frac{1}{2}m\omega^2x^2$. **7.** (a) 0.284 pm. *(b)* 2.53 keV. *(c)* 0.490 nm.
11. 1.5×10^{-15}. **13.** 0.439. **15.** 0.0527.

CHAPTER 48

Exercises

3. 9.84 kV. **5.** *(a)* 24.8 pm. *(b)* Unchanged. *(c)* Unchanged.
7. 2.1 keV. **9.** 49.6 pm; 99.2 pm; 99.2 pm. **11.** *(a)* 19.7 keV;
17.5 keV. *(b)* Zr or Nb. **15.** *(a)* $2, 0, 0, +\frac{1}{2}$.
(b) $n = 2$; $l = 1$; $m_l = 1, 0, -1$; $m_s = \pm\frac{1}{2}$.
17. Only argon would remain an inert gas. **19.** *(a)* 1.84; 2.26.
(b) 0.167; 0.119. **23.** 72 km/s^2. **25.** 5.64 cm.
27. 2.0×10^{16} s^{-1}. **29.** 3.2×10^7. **31.** 9.0×10^{-7}.
33. *(a)* None. *(b)* 240 J. **35.** *(b)* 7.33 μm. *(c)* 73.7 W/cm^2.
(d) 2.6 MW/cm^2.

Problems

1. *(a)* 5.72 keV. *(b)* 86.8 pm, 14.3 keV; 217 pm, 5.72 keV.
5. *(a)* 2.1 meV. *(b)* 2.2 nm. **7.** *(b)* − 83.3 eV. *(c)* 28.9 eV.
9. 4.74 km.

CHAPTER 49

Exercises

3. 5.90×10^{28} m^{-3}. **5.** *(a)* 0.90. *(b)* 0.69. *(c)* Sodium.
7. *(a)* 1.00; 0.986; 0.500; 0.014; zero. *(b)* 700 K. **9.** 5.53 eV.
11. 65.4 keV. **17.** 234 keV. **21.** 201°C.
23. *(a)* 5.86×10^{28} m^{-3}. *(b)* 5.51 eV. *(c)* 1390 km/s.
(d) 524 pm. **25.** *(a)* 52.1 nm. *(b)* 202. **27.** *(a)* 1.5×10^{-6}.
(b) 1.5×10^{-6}. **29.** *(a)* 5.0×10^{21} m^{-3}. *(b)* 1.7×10^5.
31. 46 nm. **33.** insulator, none, −; ext. semi., donor, n;

int. semi., none, $-$; conductor, none, $-$; conductor, none, $-$; ext. semi., acceptor, *p*. **35.** 20 GΩ; 90 Ω. **37.** (*a*) 230 nm. (*b*) Ultraviolet. **39.** Opaque.

Problems

1. 3. **3.** (*a*) 20 kJ. (*b*) 200 s. **5.** (*a*) Pure: 2.8×10^{-10}; doped: 0.034. (*b*) 0.93. **7.** (*b*) 4.9×10^{8}.

CHAPTER 50

Exercises

1. 15.7 fm. **3.** 26 MeV. **5.** 27. **7.** 13 km. **9.** 8.79 MeV. **13.** 670 keV. **15.** (*a*) 7.289 MeV. (*b*) 8.071 MeV. (*c*) $-$ 91.10 MeV. **17.** 2.53 TJ. **19.** 280 d. **21.** (*a*) 64.2 h. (*b*) 0.125. (*c*) 0.0749 **23.** (*a*) $7.57 \times 10^{16}\ \mathrm{s}^{-1}$. (*b*) $4.95 \times 10^{16}\ \mathrm{s}^{-1}$. **25.** 3.84×10^{21}. **27.** (*a*) 59.5 d. (*b*) 1.18. **29.** 87.8 mg. **33.** $Q_3 = -9.460$ MeV; $Q_4 = 4.679$ MeV; $Q_5 = -1.326$ MeV. **35.** 1.17 MeV. **37.** (*a*) 874 fm. (*b*) 6.4 fm. (*c*) No. **39.** 1.02 mg. **41.** 13 mJ. **43.** 39.4 μCi. **43.** 5.33×10^{22}. **47.** (*a*) 2.03×10^{20}. (*b*) 2.78×10^{9} Bq. (*c*) 75.1 mCi. **49.** 2.83 Gy. **51.** 1.78 mg. **53.** $-$ 1.855 MeV. **61.** (*a*) 7.19 MeV. (*b*) 12.0 MeV. (*c*) 8.69 MeV.

Problems

1. (*a*) 0.390 MeV. (*b*) 4.61 MeV. **3.** (*a*) 4. (*b*) 148 neV. (*c*) 8.38 m. (*d*) Radio region. **7.** (*a*) $3.65 \times 10^{7}\ \mathrm{s}^{-1}$. (*b*) $3.65 \times 10^{7}\ \mathrm{s}^{-1}$. (*c*) 6.41 ng. **9.** 4.268 MeV. **11.** (*b*) 960.2 keV. **13.** 730 cm^2. **15.** (*b*) 27 TW.

CHAPTER 51

Exercises

1. (*a*) 34 kg. (*b*) 12 mg. **3.** (*a*) 2.56×10^{24}. (*b*) 81.9 TJ. (*c*) 25,900 y. **9.** (*a*) 13.9 d^{-1}. (*b*) 4.97×10^{8}. **11.** $-$ 23.0 MeV. **13.** 174 MeV. **15.** 231 MeV. **17.** ${}^{238}\mathrm{U} + \mathrm{n} \rightarrow {}^{239}\mathrm{U} \rightarrow {}^{239}\mathrm{Np} + \mathrm{e}$; ${}^{239}\mathrm{Np} \rightarrow {}^{239}\mathrm{Pu} + \mathrm{e}$. **19.** 548 kg. **23.** 1.6×10^{16}. **25.** 566 W. **27.** 1.72 Gy. **31.** 450 keV. **33.** (*a*) 170 kV. **37.** (*a*) 4.0×10^{27} MeV. (*b*) 5.1×10^{26} MeV. **39.** 4.5 Gy. **41.** (*a*) 4.1 eV/atom. (*b*) 9.0 MJ/kg. (*c*) 1500 y. **45.** $K_\alpha = 3.52$ MeV; $K_\mathrm{n} = 14.07$ MeV. **47.** (*a*) 1000 km/s. (*b*) 2.0 μm.

Problems

1. (*a*) 253 MeV. **3.** (*a*) 44 kiloton. **5.** 1.4 MeV. **7.** (*b*) 2.28×10^{42} J. (*c*) 185 My. **9.** (*a*) B; $5.19N$. MeV. (*b*) A: $\frac{1}{2}N\,{}^3$H, n; B: $\frac{1}{3}N\,{}^1$H, ^{4}He, n.

CHAPTER 52

Exercises

1. (*a*) 2.4×10^{-43}. (*b*) 8.1×10^{-37}. **3.** 2.84×10^{28} m. **5.** 769 MeV. **7.** 31 nm. **9.** (*a*) K^+. (*b*) $\overline{\mathrm{n}}$. (*c*) π^0. **11.** *b*, *d*. **13.** (*a*) $\overline{\Lambda^0} \rightarrow \overline{\mathrm{p}} + \pi^+$. (*b*) $\overline{\mathrm{n}} \rightarrow \overline{\mathrm{p}} + \mathrm{e}^+ + \nu_\mathrm{e}$. (*c*) $\tau^+ \rightarrow \mu^+ + \nu_\mu + \overline{\nu_\mathrm{e}}$. (*d*) $\mathrm{K}^- \rightarrow \mu^- + \overline{\nu_\mu}$. **23.** 690 nm. **25.** (*a*) 1.6 TK. (*b*) 88 μs. **27.** (*a*) 285,000 y; era of nuclear formation. (*b*) 0.43 eV. (*c*) 0.5.

Problems

1. 31 nm. **3.** (*b*) 2.39 GK. **5.** (*a*) 280 μeV. (*b*) 4.4 mm. **7.** (*b*) $2\pi r^{3/2}(GM)^{-1/2}$.

PHOTO CREDITS

Chapter 25
Page 569: Courtesy Xerox Corporation. Page 572: GlPhotoStock/Photo Researchers. Page 583: Getty Images, Inc.

Chapter 26
Page 596 (top): Andrew Lambert Photography/Photo Researchers. Page 596 (bottom): Richard Megna/Fundamental Photographs.

Chapter 28
Page 651: SPL/Photo Researchers. Page 656: Courtesy NASA.

Chapter 30
Page 680: GlPhotoStock/Photo Researchers. Page 685: EPA/Matthias Heikel/NewsCom. Page 693: Courtesy Ken Krane.

Chapter 32
Page 729: SPL/Photo Researchers. Page 731: Courtesy Ken Krane. Page 732: Courtesy Argonne National Laboratory. Page 742: LBNL/Photo Researchers.

Chapter 33
Page 754: Sciencephotos/Alamy Limited.

Chapter 35
Page 806: Mehau Kulyk/Photo Researchers. Page 811: Courtesy R. W. De Blois. Page 812: Antony Spencer/iStockphoto.

Chapter 38
Page 866: Courtesy Stanford Linear Accelerator Laboratory. Page 873: Courtesy David W. DuBois. Page 874: Eurelios/Phototake. Page 878: Courtesy NASA.

Chapter 39
Page 884 (top): Courtesy NASA. Page 884 (bottom): SPL/Photo Researchers. Page 885 (top left): Courtesy AT&T Bell Labs. Page 885 (bottom left): Courtesy NASA. Page 885 (right): SPL/Photo Researchers. Page 887 (top): SPL/Photo Researchers. Page 887 (bottom): Brain Blake/Photo Researchers. Page 890: Photo Researchers, Inc. Page 897: Photo Researchers. Page 898: Rob Friedman/iStockphoto. Page 900: Courtesy Mount Wilson and Palomar Observatories.

Chapter 40
Page 914: Courtesy Education Development Center. Page 930: Courtesy NASA.

Chapter 41
Page 942: Courtesy Education Development Center. Page 949: Richard Megna/Fundamental Photographs. Page 952: GlPhotoStock/Photo Researchers, Inc.

Chapter 42
Page 964 (right): Ken Kay/Fundamental Photographs. Page 970: SPL/Photo Researchers. Page 971 (bottom): David Scharf/Photo Researchers.

Chapter 43
Page 989: Omikron/Photo Researchers. Pages 992 & 993 (top): Ronal R. Erickson and Museum of Holography. Page 993 (bottom): From Rigden, Physics and the Sound of Music, Scientific American, John Wiley & Sons, Inc. 1985.

Chapter 44
Page 999: Ronald Karpilo/Alamy Limited. Page 1002 (top): ©R. Mark, Experiments in Gothic Structure, MIT Press, Cambridge, MA, 1982. Page 1002 (bottom): Fundamental Photographs. Page 1003: Steve Horrell/Photo Researchers. Page 1004: From Robert Guenther, Modern Optics, John Wiley & Sons, Inc., 1990.

Chapter 46
Page 1036: Courtesy G. Mateucci. Page 1038: Courtesy A. Tonomura, J. Endo, T. Matsuda, and T. Kawasaki, Advanced Research Laboratory, Hitachi, Ltd., Kokubinju, Tokyo; H. Ezawa, Department of Physics, Gakushuin University, Mejiro, Tokyo. Page 1039: Thomas Deerinck, NCMIR/Photo Researchers; Cheryl Power/Photo Researchers. Page 1043: Courtesy Lawrence Berkeley Laboratory. Page 1048: Courtesy IBM.

Chapter 47
Page 1062: Courtesy IBM. Page 1063: From W. Finkelnburg, Structure of Matter, Springer-Verlag, 1964.

Chapter 48
Page 1090: Zephyr/Photo Researchers. Page 1092 (bottom): ©AP/Wide World Photos.

Chapter 49
Page 1119: Courtesy AT&T.

Chapter 51

Page 1154: SPL/Photo Researchers. Page 1165: Nomad/SuperStock. Page 1166 (left): Courtesy University of California, Lawrence Livermore National Laboratory. Page 1166 (right): Courtesy Los Alamos Scientific Laboratory, University of California.

Chapter 52

Page 1174: Courtesy Fermi National Accelerator Laboratory. Page 1175: ©Kyodo News International. Page 1186: American Institute of Physics/Photo Researchers. Page 1187: Courtesy AT&T. Page 1196: SPL/Photo Researchers.

INDEX

A

Absorption, photon, 1093
Accelerator(s)
  cyclotron, 731–733
  electrostatic, 651–652
Acceptor atom, 1113
Actinides, 1083
Action at a distance, 588
Adiabatic demagnetization, 809
Airy disk, 970
Alpha decay, 1136–1138
Alternating current (AC), 782–783, 845–846
  transformers of, 852–854
  and voltages, phase and amplitude relations for, 848t
Alternating current circuit
  capacitive element of, 847–848
  inductive element of, 846–847
  power factor of, 852
  power in, 851–852
  resistive element of, 846
  single-loop, 848–851
Alternator, automobile, 783
Ammeter, principle of analog, 738–740
ampere (unit), 664
  force between parallel currents, in definition of, 758
Ampère's law, 760–761
Ampèrian loop, 760–761
Amplitude reflection coefficient, 952
Amplitude transmission coefficient, 952
Analyzer, 1001–1003
Angle(s)
  of incidence, 890
  of minimum deviation, 893
  of reflection, 890, 894
  of refraction, 890, 895–896
Angular magnification, 928
Angular momentum
  of electrons in atoms, 1066–1067
  of nucleus, 1134–1135
  orbital, 1067
  spin, 1070–1071
  and uncertainty principle, 1068–1069
Angular momentum quantum number, 1067
Anisotropic materials, optically, 1004
Annihilation, 1188
Antenna, 866
Anticoincidence(s), 1023
Anticoincidence apparatus, 1023–1024
Antineutrino, 1138–1139, 1176–1177
Antineutron, 1178
Antiparticles, 570
Antiproton, 1178
Atom(s), 1062
  acceptor, 1113
  angular momentum and magnetism in, 1086–1087
  angular momentum of electrons in, 1066–1067
  Bohr's model of, 1063–1065
  donor, 1112
  excited states of, 1064, 1069–1072
    and optical transitions, 1085–1086
  fine structure of spectral lines of, 1090–1092
  ground state of, 1065–1066
  ionization energy of, 1085
  line spectrum of, 1063
  magnetism of, 1086–1087
    and atomic radiation, 1090–1092
  nuclear model of, 602–603
  nucleus of. *See* Nucleus
  photon bombardment of, 1026–1028
  quantum numbers of, 1067–1071
  shell states of, 1071
  structure of, 1062–1063
    periodic, 1083–1086
    rules of, 1082–1083
  subshell states of, 1071
  x-ray spectrum of, 1079–1081
Atomic magnetism, 805–806, 1086–1087
  and atomic radiation, 1090–1092
Atomic mass unit, energy equivalent of, 1133
Atomic number, 570, 602, 603, 1131. *See also* Quantum number(s)
  and order of elements, 1081–1082
  origin of concept of, 1081
  and periodicity of elements, 1083–1086
Atomic radiation, 1090–1092
Atomic structure, 1062–1063
  rules of, 1082–1083
Aurora borealis, 812f
Autotransformer, 858 (exercise 30)

B

Babinet's principle, 979 (problem 2)
Bainbridge's mass spectrometer, 746 (problem 5)
Balmer series, 1063
Bandwidth, 1042
Bardeen, John, 1119, 1121
Bardeen-Cooper-Schrieffer theory, 1121
Barkla, Charles, 1010
Barrier penetration, 899
Barrier tunneling, 1046–1049
  examples of, 1047–1048
  transmission coefficient in, 1047
Baryon(s), 1178
Baryon number, conservation of, 1180
Batteries, 703–704
BCS theory, 1121
Beam expanders, 938 (problem 9)
Beam splitter, 953, 1023
Beta decay, 1138–1139
Betatron, 786
Big bang cosmology
  age determination of universe in, 1192–1194
  expansion of universe in, 1186–1187
  heavy element formation in, 1191–1192
  nucleosynthesis in, 1190–1191
  particle interactions in, 1187–1190
Binding energy, 637, 1133–1134
Binding energy curve, 1133–1134
Binnig, Gerd, 1047, 1048
Bioluminescence, 886
Biot-Savart law, 753
  applications of, 753–755
Birefringence, 1005–1006
Bloch, Felix, 1089
Bohr, Niels, 1063, 1129
  correspondence principle of, 1064
  postulates of, 1063
  theory of hydrogen atom, 747 (problem 10), 1063–1065
  theory of nuclear fission, 1155–1156
Bohr frequency condition, 1063–1064
Bohr magneton, 1086, 1134
Bohr model
  of hydrogen atom, 1063–1065
  and x-ray frequencies, 1081–1082
Bohr radius, 1062
Born, Max, 1044
Bottom quark, 1184
Boundary condition, 1056
Bradley, James, 887–888, 909 (problem 1)
Bragg, W. L., 991
Bragg's law, 990–992
Brattain, Walter, 1119
Breakdown, 662, 671
Bremsstrahlung, 1080
Brewster, David, 1004

- Brewster's angle, 1003
- Brewster's law, 1004

C

- Capacitance, 679–681
  - calculating, 681–683
  - as constant of proportionality, 680
  - equivalent, 683
  - unit of, 680
- Capacitive reactance, 848
- Capacitive time constant, 714–715
- Capacitor(s), 679
  - charging, 714–715
  - cylindrical, 682–683
  - with dielectric, 687–690
  - discharging, 715–716
  - in parallel, 683–684
  - parallel plate, 681–682
  - in series, 684
  - spherical, 682
- Carbon cycle, 1172 (problem 6)
- Carnal, O., 1039
- Cavity oscillations, 865–866
- Cavity radiation, 1017
- CDF detector system, 1174
- CERN (European Center for Nuclear Physics), particle physics research at, 1175
- Cesium fountain clock, technology of, 1027
- Chadwick, James, 1154
- Charge. *See* Electric charge
- Charge density, 576
- Charged particles
  - circulating, 731–734
  - magnetism and moving, 726–727, 749–752. *See also* Magnetic field(s)
  - moving, magnetic force on, 727–731. *See also* Magnetic field(s)
- Charm quark, 1184
- Chemiluminescence, 886
- Chromatic aberration, 892, 930
- Chu, Steven, 1027
- Circuits. *See* Electric circuit(s)
- Circularly polarized wave, 1006
- Cohen-Tannoudji, Claude, 1027
- Coherent interference, 944–946
- Coherent waves, 942
- Comet tail formation, 873–874
- Complex conjugate, 1044
- Complex quantity, 1044
- Compound microscope, 929
- Compton, Arthur Holly, 1021
- Compton effect, 1021–1023
- Compton shift, 1021–1022
- Conduction band, 1111
- Conduction electrons, 571, 668, 1103–1104. *See also* Semiconductors; Superconductors
  - bands and gaps in energy states of, 1109–1111
  - density of occupied states of, 1106–1108
  - in metals, 1104–1105, 1108–1109
  - probability function of, 1105–1108
- Conductivity, 666
  - unit of, 666
- Conductors, 571–572, 661–662
  - in an electric field
    - in dynamic conditions, 663–666
    - in static conditions, 662–663
  - band structure of, 1111
  - charged, electric potential of, 649–650
  - electric field outside, 622–624
  - and Gauss' law, 621–624
- Conservation laws, of particle physics, 1179–1181
- Conservation of electric charge, 580, 702–703
- Conservation of energy, in electrostatics, 637–638
- Constructive interference, 941–942
- Contracted length, 764
- Cooper, Leon N., 1121
- Cooper pairs, 1121
- Corner reflector, 892
- Corona discharge, 650
- Correspondence principle, 1058, 1064
- Cosmic microwave background radiation, 1186–1187
- Cosmology, 1173, 1186
  - big bang, 1187–1190
    - age determination of universe in, 1192–1194
    - expansion of universe in, 1186–1187
    - heavy element formation in, 1191–1192
    - nucleosynthesis in, 1190–1191
    - particle interactions in, 1187–1190
- Coulomb, Charles A., 573
- coulomb (unit), 569
  - derivation of, 569–570
- Coulomb constant, 573–574
- Coulomb's law, 573–574
  - experimental tests of, 624–625
  - vector form of, 574–576
  - *vs.* Gauss' law, 611–612
- Critical angle, 897
- Critical temperature, 1120
- Crystalline structure
  - Bragg's law and x-ray diffraction by, 990–992
  - cubic symmetry of, 989–990
  - as diffraction grating, 988–989
  - doubly refractive, 1005–1006
  - principal indices of refraction of, 1005–1006
  - unit cell of, 990
- Cubic symmetry, 989–990
- Curie, Jacques, 1048
- Curie, Pierre, 1048
  - discovery of relation between magnetism and temperature, 809
- curie (unit), 1139–1140
- Curie constant, 809
- Curie temperature, 810
- Curie's law, 809
- Current
  - displacement, 863–864
  - electric. *See* Electric current
- Current density, 664
  - and drift speed, 664–665
- Cutoff frequency, 1020
- Cyclotron, 731–733
  - resonance condition in, 732
- Cyclotron frequency, 731
- Cylindrical capacitor, 682–683

D

- Dark matter, 1194
- Davisson, C. J., 1036
- de Broglie, Louis-Victor, 1035
- de Broglie wavelength, 1035
- de Broglie's hypothesis, testing, 1036–1038
  - Davisson-Germer experiment in, 1036–1037
  - diffraction experiment in, 1037
  - double-slit interference experiment in, 1037–1038
  - Thomson experiment in, 1036
  - Tonomura experiment in, 1037–1038
- de Haas, W. J., 1086–1087
- Debye-Scherrer experimental arrangement, 995 (question 23)
- Delayed-choice experiments, 1025–1026
- Density of states, 1105
- Depletion zone, 1115
- Destructive interference, 941–942
- Deuterium nuclei. *See* Deuteron(s)
- Deuterium-tritium fuel pellets, 1166
- Deuteron(s)
  - binding energy of, 1133–1134
  - formation of
    - in big bang nucleosynthesis, 1190–1191
    - in Sun, 1162–1163
  - fusion of, 580, 1164
- Deuteron-deuteron reaction, 1164
- Deuteron-triton reaction, 1164
- Diamagnetism, 806, 809–810
- Dichroic materials, 1005
- Dielectric constant, 671
- Dielectric materials, 671–672
  - in capacitors, 687–690
  - and Gauss' law, 688–690
- Dielectric strengths, 671
- Diffraction, 963
  - at circular aperture, 970–971
  - combined with double-slit interference, 971–975
  - Fraunhofer, 965
  - Fresnel, 965
  - multiple slit, 981–982. *See also* Diffraction grating(s)
  - single-slit, 965–967
    - intensity in, 967–970
  - vs interference, 973
  - and wave theory of light, 963–965
- Diffraction grating(s), 981–982, 985–986
  - blazed, 985
  - crystalline structure as, 988–992. *See also* x-ray diffraction
  - dispersion of, 986–987
  - intensity of lines, and slit diffraction characteristics, 990
  - making, 985
  - resolving power of, 987–988
  - secondary maxima of, 984
  - width of maxima using, 982–983
- Diffraction patterns, 963–964. *See also* Diffraction grating(s); Spectra
- Diffuse reflection, 891
- Diffusion current, 1114
- Diode lasers, 1118–1119
- Diode rectifier, 1116–1117
- Dipole
  - electric, 591
    - in electric field, 600–602
    - electric field of, 591–592
    - field lines of, 596–597, 596f
    - potential due to, 643–644
  - magnetic, 755, 801–803
  - in nonuniform field, 1088–1089
- Dipole antenna, 866–867
- Dipole moment
  - electric, 591
  - induced, 672
  - magnetic, 755
- Dirac, Paul, 815
- Direct current (DC) circuits, 701–724
- Disk of charge, uniform
  - electric field of, 593–594
  - electric potential of, 645–646
  - force on point charge, 578–579
- Dispersion
  - by grating, 986–987
  - of light, 889, 891
- Displacement current, 863–864
- Dissociation energy, 637
- Distortion parameter, 1156
- Donor atom, 1112
- Doped semiconductors, 1112–1114
- Doppler effect
  - experimental test of, 899–900
  - in light waves, 899–902
  - relativistic, 899–900
    - derivation of, 900–901

  transverse, 901–902
Double refraction, 1005–1006
Double scattering experiment, 1009–1010
Double-slit interference, 942–944
  diffraction combined with, 971–975
  intensity in, 946–948
  maxima in, 943
  minima in, 943
  wave disturbances in, adding, 947–948
  Young's experiment in, 944
Drift current, 1115
Drift speed
  and current density, 664–666

E

e-ray, 1005
Earth
  geologic record of magnetic field of, 812–813
  magnetic field of, 725–727, 811–813
  Van Allen radiation belts of, 733
Eddy currents, 780–781
Einstein, Albert
  analysis of photoelectric effect by, 1020–1021
  and concept of stimulated emission of light, 1093–1094
  on electrodynamics of moving bodies, 765
  experiment with de Haas, 1086–1087
  proof of speed of light hypothesis by, 954–955
  quantum (photon) theory of light of, 1016, 1018
Einstein-de Haas experiment, 1086–1087
Einstein's photoelectric equation, 1020–1021
Electric charge, 568
  charging by contact, 572
  charging by induction, 572
  conservation of, 580, 702–703
  continuous distributions of, 576–580
    electric field of, 592–595
    electric potential of, 644–646
  density of, 576–577
  elementary, 570
  and fluid flow, analogy of, 681
  forces of, 569, 574–576
  moving
    magnetic field of, 749–752. *See also* Magnetic field(s)
    magnetic force on, 727–731
    magnetism and, 726–727
  net, 569
  point, 573. *See also* Point charge(s)
  polarization of, 571–572
  positive and negative, 569
  quantization of, 570
  on surfaces of conductor, 621–622
  unit of measurement of, 569
Electric circuit(s)
  alternating current, 845–859. *See also* Alternating current circuit
  analysis of, 704–709
  current direction in, 705
  direct current (DC), 701
  electric current in, 701–702
  electromotive force in, 703–704
    internal resistance of, 706–709
  energy transfers in, 713
  fields in, 709–710
  inductors in, 826–827
  internal resistance of electromotive force in, 706–709
  junction rule for analyzing, 702–703
  LC, 830–833
    damped and forced oscillations in, 833–835
  loop rule for analyzing, 705
  LR, 826–827
  parallel connection of, 703
  potential difference method of analyzing, 705
  potential, differences in, 706
    path independence of, 706
  RC, 713–716
  RLC, 834–835, 848–851
Electric constant, 573
Electric current, 663, 701–703
  definition of, 664
  density of, 664
    and drift speed, 664–666
  direction of, 702
  effect of magnetic field on, 736–740
    in loop, 738–740
    in straight wire, 736–738
  and heat flow, analogy of, 667–668
  induced. *See also* Electromotive force; Magnetic flux
    eddy currents, 780–782
    Faraday's law and, 776–779
    Joule heating by, 780
    Lenz' law and, 777–778
  magnetic field of, 752–755
    in loop, 754–755
    in parallel, 756–758
    in solenoid, 758–760, 762–764
    in straight wire, 753–754, 761–762
    in toroid, 763
  SI unit of, 664
  transient behavior of, 702
Electric current loop
  magnetic dipole moment of, 802
  magnetic field of, 803
Electric dipole, 591–592. *See also* Dipole
  in electric field, 600–602
    potential energy of, 601
    torque on, 600
  field lines of, 596–597, 596f
  potential due to, 643–644
Electric dipole moment, 591
Electric dipole radiation, 867
Electric field, 588
  calculation of, 589
    from electric potential, 646–648
  in circuits, 709–710
  combined magnetic and, 730–731
  conductors in
    in dynamic conditions, 663–666
    in static conditions, 662–663
  of dipole, 591–592
  dipole in, 600–602
  electric potential calculated from, 640–641
  energy storage in, 685–687
  examples of, 589
  flux of, 613–616
  force on point charge, 589
  induced, 783–786. *See also* Magnetic flux
  insulators in, 670–672
  lines of, 595–597. *See also* Electric field lines
  nonuniform, motion in, 599–600
  outside conductor, 622–624
  point charge in, 597–598
  of point charges, 590–592
  scattering of particles by, 602–603
  sources and sinks of, 613
Electric field lines, 595–597
  conventional representation of, 595–597
  of equipotential surfaces, 648–649
  flux and, 615–616
  near conductor, 662–663
  properties of, 595
Electric flux, 613–616
Electric forces, 569, 574–576
Electric motor
  basic elements of, 783
  principle of, 738–740
Electric potential, 639–640
  calculation of, from electric field, 640–641
  of charged conductors, 649–650
  electric field calculated from, 646–648
Electric potential energy, 636–639
  of capacitor, 685–687
Electric quadrupole, 592, 606 (exercise 11), 609 (problem 4)
  potential due to, 644
Electric quadrupole moment, 644
Electromagnetic cavity oscillator, 865–866
Electromagnetic force, 1174
Electromagnetic oscillations
  damped and forced, 833–835
  forced, and resonance, 834–835
  qualitative properties of, 830–832
  quantitative properties of, 832–833
Electromagnetic oscillator, 830
Electromagnetic radiation, 886
  and Compton effect, 1021–1023
  and photoelectric effect, 1019–1021
  as thermal radiation, 1016–1019
Electromagnetic spectrum, 883–886
  gamma ray region of, 885–886
  infrared region of, 884
  microwave region of, 884
  radio wave region of, 884
  ultraviolet region of, 884–885
  visible region of, 883–884
  x ray region of, 885
Electromagnetic theory, development of, 568, 568f
Electromagnetic waves
  energy transport by, 870–872
  generation of, 866–868
  intensity of, 871–872
  Maxwell equations and, 868–870
  polarization of, 999–1013. *See also* Polarization
  pressure exerted by, 872–874
  reflection and refraction of, 891–892
  traveling, 868–870
Electromagnetism, 567–568
  displacement current in, 863–864
  and frames of reference, 764–765
  Maxwell equations of, 568, 815, 861–864. *See also* Maxwell equations
    Lorentz transformation invariance of, 765, 788
    special relativity and, 764–765
Electromotive force, 703–704
  induced, 776–777
    electric fields of, 783–786
    motional, 780–782
    potential difference and, 784–785
    practical applications of, 782–783
  internal resistance of, 706–709
  reversibility of, 704
Electron(s)
  angular momentum of, 1066–1067
  of atomic hydrogen, 1063–1072. *See also* Hydrogen, atomic

Electron(s) *(Continued)*
  bound, 1055
  conduction, 571, 1103–1104. *See also* Semiconductors; Superconductors
    bands and gaps in energy states of, 1109–1111
    density of occupied states of, 1106–1108
    in metals, 1104–1105, 1108–1109
    probability function of, 1105–1108
  configuration of, in elements, 1083–1085. *See also* Periodic table
  core, 1110
  density of occupied states of, 1106–1108
  and electric charge, 568–570. *See also* Electric charge
  energy states of, 1055–1061
    bands and gaps in, 1109–1111
  Fermi speed of, 1108–1109
  in finite potential well, 1060–1062
    energy levels of, 1060–1061
    probability densities for, 1061–1062
  free, 1055
  in infinite potential well, 1056–1060
    energy level of, 1056–1057
    probability densities for, 1057–1058
    wave numbers of, 1057–1058
  in metals, 1104–1105, 1108–1109
  and photoelectric effect, 1019–1023
  probability function of, 1105–1108
  quantized energy of, 1056. *See also* Quantum mechanics; Quantum number(s)
  wave-like behavior of, 1035–1038
  wave nature of single, 1040–1041
  wave numbers of, 1057–1058
Electron capture, 1077 (problem 10), 1138
Electron gas, 668
Electron microscope, 971, 1039–1040
Electron neutrino, 1176–1177
Electron spin, 1070–1071
Electronics, optical, 1117–1119
Electrostatic accelerator, 651–652
Electrostatic stress, 698 (problem 15)
Electrostatics, 569
  conservation of energy in, 637–638
  gravitation and, 579–580
Electroweak force, 568, 1175
Elementary charge, 570
  measuring, 598–599
Elements. *See also* Nuclide(s)
  electron configuration of, 1083–1085
  order of, and atomic number, 1081–1082
  periodic table of, 1083–1086
Equipotential surfaces, 648–649
Equipotentials, near conductor, 662–663
Esaki, Leo, 1047
Ether hypothesis, 954–955
European Center for Nuclear Physics (CERN), particle physics research at, 1175
Exchange forces, 1179
Excited state, 1057
Expansion in multipoles, 644
Extraordinary ray, 1005

F

farad (unit), 680, 775
Faraday, Michael
  capacitance concept introduced by, 680
  discovery of diamagnetism by, 809
  discovery of law of induction by, 775–776
  electric field concept introduced by, 595
  and experimentation with electric charge, 625
  and ideas on light spectra in magnetic field, 1091
  investigation of dielectric materials in capacitors by, 687
Faraday's law, of induction, 776–777
  and direction of electric current, 778–779
  induced electric fields and, 783–786
  in inertial reference frames, 786–788
  practical applications of, 782–783
  right-hand rule application in, 778–779
Farsightedness, 937 (exercise 40)
femtometer (unit), 1132
Fermat, Pierre, 894
Fermat's principle
  in derivation of law of reflection, 894–895
  in derivation of law of refraction, 896
Fermi, Enrico
  and experiments with neutron bombardment, 1154
  and theory of beta decay, 1138
fermi (unit), 1132
Fermi-Dirac probability function, 1107–1108
Fermi energy, 1106
Fermi level, 1106
Fermi National Accelerator Laboratory, particle physics discoveries at, 1184
Fermi speed, 1108–1109
Fermilab, 1184
Ferroelectric materials, 661–662
Ferromagnetism, 806, 810–811
Fiber optical devices, 897–898
Field(s), types of, 587–588
Field-effect transistors, 1119–1120
Field lines. *See* Electric field lines
Field particles, 1178–1179
Fine structure, 1090–1091
  and Zeeman effect, 1091–1092
Fission, nuclear, 1134. *See also* Nuclear fission
Fission fragments, 1154–1155
Fizeau, Hippolyte Louis, 888
Fizeau's apparatus, for measuring speed of light, 888
Fluorescence, 886
Flux, 612
  electric, 613–616
  magnetic, 776–777
  right-hand rule for sign of, 869
Flux linkages, 824
Focal length, 918
Focal point
  of lens, 924–925
  of mirror, 918
Force(s)
  basic, four, 1173–1175
  of electric charges, 569, 574–576
  electromagnetic, 1174
  electroweak, 1175
  exchange, 1179
  gravitational, 1173–1174
  strong, 1132, 1174–1175
  unification of, 1175–1176
  weak, 1174
Frames of reference. *See* Reference frame(s)
Franklin, Benjamin
  experiments of, with electric charge, 624
  positive/negative labeling of electric charge by, 568n
Fraunhofer diffraction, 965
Free charge, 689
Free-electron model, of electron conduction in metals, 668
Fresnel, Augustin, 963–964
Fresnel diffraction, 965
Fringes, interference, 942–943
  of constant thickness, 951
Frisch, Otto, 1154
Fusion, nuclear, 1134. *See also* Fusion reactor; Nuclear fusion
Fusion reactor, 1164–1167
  poloidal component of, 1164–1165
  reactions in, 1164
    inertial confinement of, 1166–1167
    magnetic confinement of, 1164–1166
    requirements for, 1164
  tokamak, 1164–1166
  toroidal field of, 1164–1166

G

Gabor, Dennis, 992
Gähler, R., 1038
Galilei, Galileo, and discussion of speed of light, 887
Galvanometer, rudiments of, 740
Gamma rays, 885–886
Gauss, Karl Friedrich, 616n
Gauss' law, 616
  application(s) of, 617–621
    to line of charge, 617–618
    to sheet of charge, 618
    to spherical shell of charge, 618–619
    to spherically symmetric charge distribution, 619–620
  and conductors, 621–624
  and Coulomb's law, 617
  and dielectric materials, 688–690
  experimental tests of, 624–625
  and flux of electric field through closed surface, 613
  for magnetism, 814–815
  and net flux through closed surface, 612
  vs. Coulomb's law, 611–612
Gaussian surface, 616
  and application of Gauss' law, 617
Geiger, Hans, 1129
Geiger counter, 632 (problem 10)
Gell-Mann, Murray, 1182
Generator
  basic elements of, 782–783
  impedance of, 854
Geometrical optics, 914. *See also* Image formation
Gerlach, Walther, 1087
Germer, L. H., 1036
Giaver, Ivar, 1047
Gilbert, William, 811
Glashow, Sheldon, 1175
Gluon(s), 1183–1184
Goudsmit, Samuel, 1070
Grand unified theories (GUTs), 1175
Grating spectroscope, 985
Gravitation, and electrostatics, 579–580
Gravitational field, Earth's, 587–588
Gravitational force, 1173–1174. *See also* Gravity
Ground state, 1057
Grounded object, 571

H
Hahn, Otto, 1154
Hall, Edwin, 734
Hall effect, 734–736
  quantized, 736
Hall potential difference (Hall voltage), 735
Hall resistance, 736
Heat flow, electric current and, analogy of, 667–668
Heaviside, Oliver, 861
Heavy element formation, in big bang cosmology, 1191–1192
Heisenberg, Werner, 1042
Heisenberg's uncertainty principle, 1042–1043
Helmholtz coil, 772 (problem 1)
Henry, Joseph, 775, 823
henry (unit), 775, 823
Hertz, Heinrich, 864, 870
Holography, 992–993
Homopolar generator, 798 (problem 8)
Hubble, Edwin, 1186
Hubble parameter, 1186
Hull, G. F., 873
Huygens, Christiaan, 893
  studies of double refraction, 1005
Huygens' principle
  in derivation of law of reflection, 893–894
  in derivation of law of refraction, 895–896
Hydrogen, atomic
  Bohr model of, 1063–1065
  energy levels of, 1064
  excited states of, 1069–1072
  ground state of, 1065–1066
  line spectrum of, 1063
  quantum states of, 1071–1072
  radial probability density of, 1066
  2*p* state of, 1069–1070
  2*s* state of, 1069
Hysteresis, 810
  and magnetic domains, 810–811
Hysteresis curve, 810f

I
Image, 913
Image formation, 913–914
  by lenses, 923–928
  mirror equations in, 918
  by optical instruments, 928–930
  by plane mirror, 914–917
  ray tracing in, 919–920
  real and virtual, 918–919
  by spherical mirror, 917–921
  by spherical refracting surfaces, 921–923
Impedance
  of generator, 854
  of *RLC* circuit, 849
Incandescence, 886
Incidence
  angle of, 890
  plane of, 890
Incident ray, 890
Incoherent interference, 945–946
Incoherent waves, 942
Index (indices) of refraction, 890
  of doubly refracting crystals, 1005
  principal, 1005
Induced dipole moment, 672
Induced electric current, 776
Induced electromotive force, 776
Induced surface charge, 670, 689
Inductance, 823–824
  calculating, 824–826
  and energy density in magnetic field, 829–830
  and energy storage in magnetic field, 827–829
  in *LC* circuits, 830–835
  in *LR* circuits, 826–827
  of solenoid, 824–825
  of toroid, 825
Induction, charging by, 572
Induction furnace, 781
Inductive reactance, 847
Inductive time constant, 826
Inductor(s), 823
  flux linkages of, 824
  with magnetic materials, 825–826
Infinite well, 1056
Infrared rays, 884
Insulators, 571, 661–662
  band structure of, 1111
  in electric field, 670–672
Intensity, wave
  in double-slit interference, 946–948
  in single-slit diffraction, 967–970
Interference, 941
  coherent, 944–946
  constructive, 941–942
  destructive, 941–942
  double-slit, 942–944
    diffraction combined with, 971–975
    intensity in, 946–948
    maxima, 943
    minima, 943
  fringes caused by, 942–943
  incoherent, 945–946
  maxima in, 942–943
  and Michelson's interferometer, 953–955
  minima in, 942–943
  multiple slit, 981–984
  from thin films, 948–952
  vs diffraction, 973
Interferometer, 953–954
International Thermonuclear Experimental Reactor (ITER), 1165
Interplanar spacings, 990–991
Ion(s), 662
Ionization, 662
Ionization energy, 637, 1085
Ionizing radiation, 1139
  units of, 1139–1140
Isotopes, 1131
Isotropic materials, optically, 1004
Ives, H. E., 899–900

J
Josephson, Brian, 1047
Joule heating, 713
  by induced electric current, 780
Junction rule, 702–703
Junction transistors, 1119

K
Kirchhoff, Gustav, 894, 1017
Kirchhoff's first law, 702–703
Kirchhoff's second law, 705
Klystron, 866
Kundig, Water, 901

L
Lamb, W. E., 1020
Lanthanides, 1083
Laser(s), 1092–1096
  diode, 1118–1119
Laser cooling, 1027–1028
Laser fusion, 1166–1167
Laser light
  Einstein's concept of, 1093–1094
  production of, 1094–1096
  properties of, 1092–1093
Laue spots, 989
Law of Malus, 1002
Law of reflection, 890
  derivation of, 893–895
Law of refraction, 890
  derivation of, 895–897
Lawrence Livermore National Laboratory, 1166–1167
Lawson's criterion, 1164
LC circuit, 830–833
  frequency of oscillations of, 832–833
  natural frequency of, 834
  oscillating, and analogy to simple harmonic motion, 832
  resonance condition of, 835
Lens(es)
  converging, 924
  and derivation of thin-lens formulas, 925–926
  diverging, 924
  first focal point of, 924
  image formation by, 923–928
    inverted, 924
    ray tracing in, 925
    sign convention used in, 924–925
  second focal point of, 924–925
  two, image formation by, 926–928
Lens maker's equation, 923
Lenz, Heinrich, 777–779
Lenz' law, 777–779
Lepton(s), 1176–1177
Lepton number, conservation of, 1179–1180
Libby, Willard, 1141
Light, 883–884. *See also* Electromagnetic radiation; Electromagnetic waves
  diffraction and wave theory of, 963–965. *See also* Diffraction; Interference
  Doppler effect of, 899–902
  nature of
    double slit experiment and, 1016
    experimental studies of, 1023–1028
    particle-like, 1015–1016, 1025
    wave, 1015, 1024–1025
  photon theory of, 886, 1016–1019. *See also* Photon(s)
  polarized, 999–1013. *See also* Polarization
    use in industries, 1002–1003
  propagation of, in matter, 889
  quantum theory of, 1018
  reflection of, 890–891. *See also* Reflection
  refraction of, 890–891. *See also* Refraction
  spectrum of white, 986
  speed of, 887–890
    in matter, 889
  visible, 886–887
  wave theory of, 963–965
Light-emitting diodes, 1117–1118
Line of charge, uniform
  electric field of, 592–593
  electric potential of, 644–645
  force on point charge in, 577
  Gauss' law applied to, 617–618
Linear charge density, 576
Linear materials, 671
Liquid-drop fission model, 1155–1156
Lloyd's mirror experiment, 953
Loop rule, 705
Lorentz force, 730–731
Lorentz transformation, and invariance of laws of electromagnetism, 765, 788
LR circuits, 826–827
Luminescence, 886
  types of, 886–887

M
Magnetic braking, 781
Magnetic constant, 751
Magnetic damping, 793 (question 25)
Magnetic dipole, 755, 801–803
  forces on, in nonuniform field, 804–805

Magnetic dipole *(Continued)*
  magnetic field of, 802–803
Magnetic dipole moment, 755
  of current loop, 802
  induced, 804–805
  selected values of, 803t
Magnetic domains, 811
  hysteresis and, 810–811
Magnetic field
  Ampère's law in calculation of, 760–761
    applications of, 761–764
  Biot-Savart law in calculation of, 753
    applications of, 753–755
  combined electric and, 730–731
  of dipole, 803
  effect on electric current, 736–740
    in loop, 738–740
    in straight wire, 736–738
  of electric current, 752–755
    in loop, 754–755
    in parallel, 756–758
    in solenoid, 758–760, 762–764
    in straight wire, 753–754, 761–762
    in toroid, 763
  energy density of, 829–830
  energy storage in, 827–829
  Hall effect of, 734–736
  of moving charged particles, 749–752
  nonuniform
    motion in, 734
    particle entrapment in, 733–734
  typical values of, 729
Magnetic flux, 776–777
  and eddy currents, 780–781
  electric fields induced by, 783–786
  Faraday's law and Lenz' law for, 777–779
  unit of, 776
Magnetic force, on moving charged particle, 727–731
Magnetic lenses, 1039
Magnetic materials, 808–811
  in inductors, 825–826
Magnetic mirror, 733–734
Magnetic monopoles, 726, 815
  and Maxwell equations, 864
Magnetic poles, 727
  rule of interaction of, 727
Magnetic quantum number, 1067
Magnetic resonance imaging (MRI), 806, 1090
Magnetism
  atomic, 805–806, 1086–1087
  field vectors in, 727
  and moving charged particles, 726–727
  nuclear, 806–807
  of planets, 811–814
Magnetization, 807–808
  saturation value of, 809
Magnetostatics, 751–752
Magnetron, 866
Majority carriers, 1112
Malus, Etienne Louis, 1002
Marsden, Ernest, 1129
Mass excess, 1148 (exercise 15)
Matter waves, 1035–1036, 1041–1042
  applications of, 1039–1040
  electrons as, 1035–1038. *See also* Electron(s)
  frequency of, 1043–1044
  Heisenberg's uncertainty relationships and, 1042–1043
  particle, 1038–1039
  Schrödinger's equation and, 1045–1046
  wave function of, 1044–1045
  wave number of, 1042–1043
Maxima, interference, 942, 943
  principal, 982–983
  secondary, 982, 984
Maxwell, James Clerk, 861
  and generalization of Ampère's law, 862
Maxwell equations
  Ampère's law, 760, 861–862, 865t
  as basics of electromagnetism, 861–862
  consistency of, with special theory of relativity, 864
  electromagnetic wave prediction from, 864
  electromagnetism and, 862–864
  Faraday's law, 784, 861–862, 865t
  Gauss' law for electricity, 861–862, 865t
  Gauss' law for magnetism, 815, 861–862, 865t
  magnetic monopoles and, 864
  speed of light from electromagnetic considerations by, 870
  symmetry feature of, 864
  traveling waves and, 868–870
megaparsec (unit), 1186
Meitner, Lise, 1154
Mendeleev, D., 1081
Mesons, 1177–1178
Metal-oxide-semiconductor FET, 1120
Metals
  band structure of, 1111
  conduction electrons in, 1104–1105, 1108–1109
    bands and gaps in energy states of, 1109–1111
    density of occupied states of, 1106–1108
    filling allowed states of, 1105–1108
  energy bands and band structure of, 1109–1114
  work function of, 1109–1111
Metastable, 1093
Michelson, Albert A., 888, 953
Michelson-Morley experiment, 954–955
Michelson's interferometer, 953–954
  and light propagation, 954–955
microfarad (unit), 680
Microscope(s)
  compound, 929
  diffraction reduction in, 971
  electron, 971, 1039–1040
  scanning electron, 1039
  scanning tunneling, 1047–1048
Microwaves, 884
Millikan, Robert A., 598, 1032
Millikan oil-drop apparatus, 598–599
Minima, interference, 942, 943
Minimum energy principle, 1083
Minority carriers, 1112
Mirror equations, 918
  derivation, 920–921
Mirrors
  plane, 914–917
  spherical, 917–921
Mlynek, J., 1039
Monopole(s), 644
  magnetic, 726, 815
    and Maxwell equations, 864
Morley, E. W., 954
Moseley, H. G. J., 1081
Moseley plot, 1081–1082
Motional electromotive force, 780–782
Multiple slit interference, 981–984. *See also* Diffraction grating(s)
Muon neutrino, 1177

N

n-type semiconductors, 665, 1112
National Ignition Facility, 1167
Natural frequency, of *RLC* circuit, 834
Natural radioactivity, 1140–1141
Near field, 867
Near point, 928
Nearsightedness, 937 (exercise 40)
Nebulae, 1186
Neutrino(s), 1138–1139, 1176–1177
Neutrino oscillation, 1177
Neutron(s)
  diffraction of, 1038–1039
  in nucleus, 1131
  thermal, 1040
Neutron number, 1131
Newton's rings, 952–953
Nichols, E. F., 873
Nonpolar dielectrics, 671
Normalization equation, 1058
NOVA laser fusion project, 685, 1166
Nuclear fission, 1134, 1153, 1157–1159. *See also* Nuclear reactors
  basic process of, 1154–1155
  potential barrier, 1156
  theory of, 1155–1157
Nuclear fusion, 1134, 1153. *See also* Thermonuclear fusion
Nuclear magnetic moment, 1134
Nuclear magnetic resonance, 806, 1089–1090
Nuclear magnetism, 806–807, 1134–1135
Nuclear magnetron, 1134
Nuclear reactions, 1141–1143. *See also* Nuclear reactors
  elastic scattering in, 1142
  endothermic, 1142
  exothermic, 1142
  inelastic scattering in, 1142
  threshold energy of, 1142
Nuclear reactors
  accumulation of radioactive waste of, 1159
  basic principles of, 1157–1159
  chain reaction in, 1157
  control rods of, 1158
  critical operation of, 1158
  multiplication factor parameter of, 1157–1158
  natural, 1159–1161
  neutron capture problem of, 1157
  neutron energy problem of, 1157
  neutron leakage problem of, 1157
  pressurized-water, 1158
  response time and power level of, 1158
Nuclear spin, 1134–1135
Nucleon(s), 1131
Nucleosynthesis, 1190–1191
Nucleus, 602
  angular momentum of, 1134–1135
  collective model of, 1144
  discovery of, 1129–1131
  electric charge of, 570, 602–603
  forces within, 1132
  independent-particle model of, 1144–1145
  mass of, and binding energies, 1133–1134
  models of, 1143–1145
  neutrons in, 1131
  nucleons of, 1131
  particle reactions in, 1141–1143
  protons in, 1131
  radioactive decay of, 1135–1136
    ionizing radiation from, 1139–1140
  radius of, 1132–1133
    determining, 603
Nuclide(s), 1131
  expanded chart of, 1142
  plot of known, 1131
  properties of selected, 1132

O

- *o*-ray, 1005
- Oersted, Hans Christian, 749
- Ohmic materials, 666–668
- Ohm's law, 666–667
  - at microscopic level, 668–670
- Onnes, Kammerlingh, discovery of superconductivity by, 1120
- Optic axis, 1005
- Optical activity, 1008
- Optical electronics, 1117–1119
- Optical instrument(s), 928
  - angular magnification of, 928
  - compound microscope as, 929
  - refracting telescope as, 929–930
  - simple magnifier as, 928–929
- Optical path length, 896, 925
- Optical reversibility, and phase changes on reflection, 952–953
- Optics
  - adaptive, 930
  - fiber, 897–898
  - geometrical, 914. *See also* Image formation
  - physical, 914. *See also* Physical optics
  - ray, 914
  - wave, 914
- Orbital angular momentum, 1067
- Ordinary ray, 1005
- Oscillator
  - electromagnetic, 830
  - electromagnetic cavity, 865–866

P

- *p*-type semiconductors, 1112
- Pair production, 1188
- Pairing energy, 1121
- Pairing gap, 1121
- Parallel capacitors, 683–684
- Parallel plate capacitors, 681–682
- Paramagnetism, 806, 808–809
- Paraxial rays, 918, 921
- Particle(s). *See also* Particle physics
  - charged. *See* Charged particles; Electric charge
  - classification of, 1176–1179
  - elementary, 1184–1185. *See also* Particle physics
  - field, 1178–1179
  - quantum mechanics of, 1035–1046. *See also* Electron(s); Matter waves
- Particle accelerators
  - betatron, 786
  - cyclotron, 731–733
  - synchrotron, 733
- Particle physics, 1173–1186
  - conservation laws of, 1179–1181
  - interactions in, 1173–1176
  - particle classification in, 1176–1179
  - quark model of, 1181–1184
  - Standard Model of, 1184–1185
  - terminology used in, 1184
- Pauli, Wolfgang, 1083, 1087
- Pauli exclusion principle, 1083
- Penzias, Arno, 1187
- Periodic table, 1083–1086
- Permeability, 807
  - of diamagnetic substances, 810t
  - of paramagnetic materials, 808–809, 808t
- Permeability constant, 751
- Permittivity
  - of free space, 573
  - of materials, 672
- Phasor diagram, 846
- Phillips, William D., 1027
- Phosphor, 886
- Phosphorescence, 886
- Photoelectric effect, 1019–1020
- Photon(s), 1016, 1024–1025
  - absorption of, 1093
  - and Compton effect analysis, 1021–1023
  - concept of, 1018
    - experimental evidence supporting, 1023–1028
    - theoretical development of, 1016–1019
  - designated, 1023
  - and photoelectric effect, 1019–1020
  - and photoelectric effect analysis, 1020–1021
  - trigger, 1023
- Photon bombardment, 1026–1028
- Physical optics, 914
  - diffraction in, 963–980. *See also* Diffraction
  - wave interference in, 941–961
- picofarad (unit), 680
- Piezoelectricity, 1048
- Pion(s), 1177–1178
- Planck, Max, discovery of quantization of energy, 1016
- Planck constant, 1016
- Planck's radiation law, 1017–1019, 1187
- Plane mirror
  - image formation in, 914–917
  - image of extended object in, 915–916
  - image reversal in, 916–917
- Plane of incidence, 890
- Plane polarization, 1000
- Planets, magnetic field of, 811–814, 813t
- Plasma(s), 733
- Plimpton and Lawton, experiments of, with electric charge, 625
- *pn* junction semiconductors, 1114–1116
- Point charge(s), 573
  - collection of
    - potential due to, 642–643
  - in electric field, 597–598
  - electric fields of, 590–592
  - potential due to, 641–644
  - potential energy of system of, 638–639
- Point source, 914 (footnote)
- Poisson, Simeon-Denis, 964
- Poisson spot, 964
- Polarization, 670
  - of electric charges, 571–572
  - of electromagnetic waves, 999–1001
    - analyzers and polarizers in, 1001–1003
    - circular, 1006–1008
    - direction of, 1000
    - and double refraction, 1004–1006
    - industrial use of, 1002–1003
    - linear, 1000
    - optical activity and, 1008
    - plane of, 1000
    - by reflection, 1003–1004
    - by scattering, 1008–1010
    - sheets causing, 1001–1003
- Polarizers, 1001–1003
- Polarizing angle, 1003
- Polarizing sheets, 1001–1003
- Polaroid, 1001
- Population inversion, 1094
- Positron, charge of, 570
- Potential energy, 635–636
  - electric, 636–639
- Potential energy barrier, 1046
- Potential well(s), 1055–1056
  - finite, 1060–1061
  - infinite, 1056–1060
    - energy level of, 1056–1057
    - probability densities for, 1057–1058
- Poynting, John Henry, 870
- Poynting vector, 870–872
- Pressurized-water reactors, 1158
- Priestly, Joseph, experiments of, with electric charge, 625
- Primary loop, 1158
- Princeton Plasma Physics Laboratory, 1165
- Principal indices, of refraction, 1005–1006
- Principal maxima, 982–983
- Principal quantum number, 1067
- Prism spectrograph, 986
- Probability density, 1044
- Probability function, 1105–1108
  - Fermi-Dirac, 1107–1108
- Proper length, 764
- Proper time, 900
- Proton(s), 1131
- Proton number, 1131
- Proton-proton cycle, 1162–1163
- Purcell, Edward, 1089

Q

- *Q* factor, 843 (problem 15)
- *Q*-value, of radioactive decay, 1136
- Quadrupole, electric, 592, 606 (exercise 11), 609 (problem 4)
  - potential due to, 644
- Quantum corral, 1062
- Quantum electrodynamics, 568
- Quantum mechanics. *See also* Matter waves
  - and barrier tunneling, 1046–1049
  - initial development of, 1018
  - matter waves and, 1035–1036, 1041–1042
    - applications of, 1039–1040
    - of electrons, 1035–1038
    - Heisenberg's uncertainty relationships in, 1042–1043
    - of particles, 1038–1039
    - properties of, 1042–1045
    - Schrödinger's equation in, 1045–1046
  - and periodicity of elements, 1083–1086
  - and wave nature of electrons, 1040–1041. *See also* Electron(s)
- Quantum number(s)
  - angular momentum, 1067
  - atomic, 1067–1069
    - and periodicity of elements, 1083–1086
  - magnetic, 1067
  - principal, 1067
  - spin magnetic, 1071
  - total angular momentum, 1091
- Quantum number principle, 1083
- Quantum state, concept of, 1063–1064
- Quantum statistical mechanics, 1103–1104
- Quark(s), 570, 1181–1184
- Quark model, 1181–1184
  - interaction between quarks in, 1183–1184
- Quarter-wave plate, 1006

R

- r-process, 1192
- rad (unit), 1140
- Radial probability density, 1066
- Radiation
  - cavity, 1017
  - electric dipole, 867
  - electromagnetic, 867–868
  - ionizing, 1139–1140
  - thermal, 886, 1017
    - experimental investigation of, 1016–1019
    - properties of, 1017–1019
- Radiation field, electromagnetic, 867–868
- Radiation pressure, 872–874
  - and formation of comet tail, 873–874
  - measurement of, 873
- Radio waves, 884

Radioactive dating, 1141
Radioactive decay, 1135–1136
  alpha decay, 1136–1138
  beta decay, 1138–1139
  Q-value of, 1136
Radioactive isotopes, 1131, 1141
Radioactivity, natural, 1140–1141
Radiocarbon dating, 1141
Radionuclides, 1131
Radon, 1140
Rare earths, 1083
Ray optics, 914
Ray tracing
  in lens images, 925
  in spherical mirror images, 919–920
Rayleigh's criterion, 970
RC circuits, 713–716
Real image, 913, 918
Red giant phase, of star evolution, 1191–1192
Red shift, 900
Reference frame(s), electromagnetism and, 764–765, 786–788
Reflected ray, 890
Reflection
  angle of, 890, 894
  diffuse, 891
  of electromagnetic waves, 891–892
  law of, 890, 893–895
  of light, 890–891
  polarization by, 1003–1004
  total internal, 892, 897–899
Reflection gratings, 985
Refracted ray, 890
Refracting telescope, 929–930
  adaptive optics for, 930
  chromatic aberration of, 930
  field of view of, 930
  Hubble, 930
  light-gathering power of, 930
  spherical aberration of, 930
Refraction
  angle of, 890, 895–896
  double, 1004–1006
  of electromagnetic waves, 891–892
  index of, 890–891
  law of, 890, 895–897
  of light, 890–891
  from spherical surfaces, 921–923
Reid, A., 1036
Relativistic Doppler effect, 899–900
  derivation of, 900–901
rem (unit), 1140
Resistivity, 666
  of semiconductors, 1112
  temperature variation of, 668
Resistors
  in parallel, 710–711
  in series, 711
Resolving power, of grating, 987–988
Resonance capture, 1157
Resonance condition
  of *RLC* circuit, 835
Reverse-biased connections, 1116–1117
Reverse saturation current, 1127 (problem 7)
Right-hand rule
  applied in Faraday's law of induction, 778–779
  for sign of flux, 869
Ring of charge, uniform
  electric field of, 593–594
  electric potential of, 645
  force on point charge, 577–578
Ritz combination principle, 1075 (exercise 25)
RLC circuit, 834–835, 848–851
  damped and forced oscillations in, 833–835
  differential analysis of, 850–851
  graphical analysis of, 849–850
  impedance of, 849
  natural frequency of, 834
  power in, 852
  resonance condition of, 835
  single loop, 848–851
  trigonometric analysis of, 849
Roemer, Ole, 887, 909 (problem 2)
roentgen (unit), 1140
Rohrer, Heinrich, 1047, 1048
Rowland, Henry, 749
Rubbia, Carlo, 1175
Ruska, Ernst, 1039, 1048
Rutherford, Ernest
  alpha particle scattering experiments of, 602–603
  discovery of atomic nucleus by, 1129–1131
  as founder of atomic and nuclear physics, 603

S

s-process, 1192
Salam, Abdus, 1175
Scalar field, 587
Scanning electron microscope, 1039
Scanning tunneling microscope, 1047–1048
Scattering, of light, 1008–1009
  double, 1009–1010
Schrieffer, J. Robert, 1121
Schrödinger, Erwin, 1044, 1045
Schrödinger's equation, 1045–1046
  applied to barrier tunneling, 1046–1049
  applied to free particle, 1045–1046
Scully, M. O., 1020
Secondary loop, 1158
Secondary maxima, 982, 984
Secular equilibrium, 1150 (problem 5)
Selection rules, 1085
Semiconductors, 662, 1111
  band structure of, 1111
  density of charge carriers in, 1111–1112
  devices using, 1114–1116
  in diode lasers, 1118–1119
  in diode rectifier, 1116–1117
  doped, 1112–1114
  extrinsic, 1112
  intrinsic, 1112
  n-type, 665, 1112
  in optical electronics, 1117–1119
  p-type, 1113
  pn junction, 1114–1116
  resistivity of, 1112
  temperature coefficient of resistivity of, 1112
  in transistors, 1119–1120
Sheet of charge, infinite/uniform
  electric field of, 594
  field lines of, 595, 596f
  Gauss' law applied to, 618
Shell, atomic, 1071
Shell of charge, uniform spherical, electric field of, 594–595
Shell theorems, electrostatic, 579
Shockley, William, 1119
siemens (unit), 666
sievert (unit), 1140
Simple magnifier, 928–929
Single-slit diffraction, 965–967
  intensity in, 967–970
Snell's law, 890
Sodium doublet, 1091
Solar compass, 1009
Solenoid, 758
  inductance of, 824–825
  magnetic field of, 758–760
Solids, electrical conductivity of, 1103–1104
Space quantization, 1067, 1087–1089
Special theory of relativity
  in development of electromagnetic theory, 568
  and reference frames in electromagnetism, 764–765, 786–788
Spectra, 985–986
Spectral lines, 985–986
Spectral radiancy, 1017
Spectrograph(s), 985–986
  prism, 986
Speed of light, 887–890
  historical measurement of, 887–888, 888t
  in matter, 889, 889t
Spherical aberration, 930
Spherical capacitor, 682
Spherical mirror
  focal length of, 918
  focal point of, 918
  image formation in, 917–918
  mirror equation derivation for, 920–921
  real image formed by, 918
  sign convention use with, 918–919
  virtual image formed by, 918
Spherical refracting surfaces, 921–923
Spherical shell of charge, uniform
  electric field of, 594–595
  force on point charge in, 579–580
  Gauss' law applied to, 618–619
Spin angular momentum, 1070–1071
Spin flip, 1089–1090
Spin magnetic quantum number, 1071
Spontaneous emission, 1093
Standard Model, of elementary particles, 1184–1185
Stanford Linear Accelerator, 866f
Stars
  red giant phase of, 1191–1192
  thermonuclear fusion in, 1162–1163
Static field, 587
Stefan, Josef, 1017
Stefan-Boltzmann constant, 1017–1019
Stefan-Boltzmann law, 1017–1019
Stern, Otto, 1087
Stern-Gerlach experiment, 1087–1089
Stilwell, G. R., 899–900
Stimulated emission, 1093–1094, 1094
Stopping potential, 1019
Strangeness, 1181
Strassmann, Fritz, 1154
Strong force, 637–638, 1132, 1174–1175
Subshell, atomic, 1071
Sun, thermonuclear fusion in, 1162–1163
Super-Kamiokande detector, 1175
Supercaps, 683
Superconductors, 662, 1120–1122
Superposition, principle of
  applied to electric field of point charges, 590
  applied to electric forces, 575
Surface charge, induced, 689
Surface charge density, 576
Symmetry, cubic, 989–990
Synchrotron, 733

T

Temperature, critical, 1120
tesla (unit), 729
Theory of everything (TOE), 1175
Thermal neutrons, 1040, 1154–1155
Thermal radiation, 886, 1017
  experimental investigation of, 1016–1019
  properties of, 1017–1019
Thermonuclear fusion, 1161–1162

controlled, 1164–1167. *See also* Fusion reactor
in stars, 1162–1163
Thin film interference, 948–952
equations for, 950–952
fringes of constant thickness in, 951
optical path difference *vs.* geometrical path difference in, 950
optical reversibility and phase changes on reflection, 952–953
phase change on reflection in, 949–950
Thin-lens approximation, 926
Thin-lens formula derivation, 925–926
Thomson, George P., 1036
Thomson, J. J., 602
atomic model of, 1130
Thomson model, 602, 1130
Time-varying field, 587
Tokamak, 1164–1166
Tokamak Fusion Test Reactor, 1165
Tonomura, A., 1037–1038
Top quark, 1184
Toroid, inductance of, 825
Torsion balance, 573
Total angular momentum quantum number, 1091
Total internal reflection, 892, 897–899
frustrated, 898–899
Transformer, 852–854
step, 853
Transients, 846
Transistors, 1119–1120
field-effect, 1119–1120
junction, 1119
Transmission coefficient, 1047
Transmission line, 866
Transverse Doppler effect, 901–902
Triboluminescence, 886–887
Tritium, 580
Triton, 1164
Tunnel diode, 1047

U

Uhlenbeck, George, 1070
Ultraviolet rays, 884–885
Uncertainty principle, angular momentum and, 1068–1069
Unit cell, 990
Universe
age determination of, 1192–1194
big bang theory of formation of, 1187–1192
expansion of, 1186–1187

V

Valence band, 1111
Van Allen radiation belts, 733
Van de Graaff, Robert J., 651
Van de Graaff accelerator, 651–652, 1141
van der Meer, Simon, 1175
Vector field, 587
flux of, 612–613
Velocity selector, 730–731
Virtual image, 913
Virtual object, 919
Voltage, 639
Voltage divider, 706
Voltmeter, principle of analog, 740
Volume charge density, 576
von Klitzing, Klaus, 736
von Klitzing constant, 736
von Laue, Max, 989

W

Water, polar dielectric properties of, 671–672
Wave(s)
coherent, 942
diffraction of, 963–980. *See also* Diffraction
electromagnetic, 866–868. *See also* Electromagnetic waves; Light
incoherent, 942
interference of. *See* Interference
light, 883–884. *See also* Electromagnetic radiation; Light
matter, 1035–1036, 1041–1042. *See also* Matter waves
Wave function, 1044–1045
Wave number, 1041
Wave optics, 914, 941. *See also* Diffraction; Interference
Wave packet, localizing
in space, 1041–1042
in time, 1042
Wavetrain, finite, 945
Weak force, 1174
weber (unit), 776
Weinberg, Stephen, 1175
Wheatstone bridge, 723 (problem 12)
Wheeler, John, theory of nuclear fission, 1155–1156
Wien, Wilhelm, 1017
Wien's displacement law, 1017
Wiles, Andrew, 894
Wilson, Robert, 1187
Work function, 1020, 1109–1111

X

x-ray(s), 885, 988–989
and numbering of elements, 1080
x-ray diffraction, 988–989
Bragg's law and, 990–992
by crystalline structures, 989–992
interplanar spacings in, 990–991
molecular structure revealed by, 991
x-ray spectrum
of atoms, 1079–1081
characteristic, 1080
continuous, 1079–1080

Y

Young, Thomas, 944
polarization studies of, 1002
Young's double-slit experiment, 944

Z

Zeeman, Pieter, 1091
Zeeman effect, 1091–1092
Zeilinger, A., 1038
Zweig, George, 1182

SOME PHYSICAL CONSTANTS*

Speed of light in vacuum	c	3.00×10^{8} m/s
Newtonian gravitational constant	G	6.67×10^{-11} N · m^2/kg^2
Avogadro constant	N_A	6.02×10^{23} mol^{-1}
Molar gas constant	R	8.31 J/mol · K
Mass-energy relation	c^2	8.99×10^{16} J/kg
		931.5 MeV/u
Electric constant (permittivity)	ϵ_0	8.85×10^{-12} F/m
Magnetic constant (permeability)	μ_0	1.26×10^{-6} H/m
Planck constant	h	6.63×10^{-34} J · s
		4.14×10^{-15} eV · s
Boltzman constant	k	1.38×10^{-23} J/K
		8.62×10^{-5} eV/K
Elementary charge	e	1.60×10^{-19} C
Electron mass	m_e	9.11×10^{-31} kg
Electron rest energy	$m_e c^2$	511.0 keV
Proton mass	m_p	1.67×10^{-27} kg
Proton rest energy	$m_p c^2$	938.3 MeV
Bohr radius	a_0	5.29×10^{-11} m
Bohr magneton	μ_B	9.27×10^{-24} J/T
		5.79×10^{-5} eV/T

*For a more complete list, showing also the best experimental values, see Appendix B.

SOME CONVERSION FACTORS*

Mass

1 kg = 1000 g = 6.02×10^{26} u
1 u = 1.66×10^{-27} kg

Length

1 m = 100 cm = 39.4 in. = 3.28 ft
1 mi = 1.61 km = 5280 ft
1 in. = 2.54 cm
1 light-year = 3.26 parsec = 9.46×10^{15} m
1 Å = 0.1 nm = 100 pm = 10^{-10} m

Time

1 d = 86,400 s
1 y = $365\frac{1}{4}$ d = 3.16×10^{7} s

Volume

1 L = 1000 cm^3 = 10^{-3} m^3 = 1.06 quart
1 gal (U.S.) = 231 in.^{3} = 3.79 L

Angular measure

1 rad = 57.3° = 0.159 rev
π rad = 180° = $\frac{1}{2}$ rev

Speed

1 m/s = 3.28 ft/s = 2.24 mi/h
1 km/h = 0.621 mi/h

Force and Pressure

1 N = 10^5 dyne = 0.225 lb
1 Pa = 1 N/m^2 = 10 dyne/cm^2 = 1.45×10^{-4} lb/in.^{2}
1 atm = 1.01×10^5 Pa = 14.7 lb/in.^{2} = 76 cm-Hg

Energy and Power

1 J = 10^7 erg = 0.239 cal = 0.738 ft · lb
1 kW · h = 3.6×10^6 J
1 cal = 4.19 J
1 eV = 1.60×10^{-19} J
1 horsepower = 746 W = 550 ft · lb/s

Electricity and magnetism

1 T = 1 Wb/m^2 = 10^4 gauss

*See Appendix G for a more complete list.

Rh __

Text__

SOME PHYSICAL PROPERTIES

Air (dry, at 20 °C and 1 atm)	
Density	1.21 kg/m^3
Specific heat capacity at constant pressure	1010 J/kg · K
Ratio of specific heat capacities	1.40
Speed of sound	343 m/s
Electrical breakdown strength	3×10^6 V/m
Effective molar mass	0.0289 kg/mol
Water	
Density	1000 kg/m^3
Speed of sound	1460 m/s
Specific heat capacity at constant pressure	4190 J/kg · K
Heat of fusion (0 °C)	333 kJ/kg
Heat of vaporization (100 °C)	2260 kJ/kg
Index of refraction ($\lambda = 589$ nm)	1.33
Molar mass	0.0180 kg/mol
Earth	
Mass	5.98×10^{24} kg
Mean radius	6.37×10^6 m
Free fall acceleration at the Earth's surface	9.81 m/s^2
Standard atmosphere	1.01×10^5 Pa
Period of satellite at 100 km altitude	86.3 min
Radius of the geosynchronous orbit	42,200 km
Escape speed	11.2 km/s
Magnetic dipole moment	8.0×10^{22} A · m^2
Mean electric field at surface	150 V/m, down
Distance to:	
Moon	3.82×10^8 m
Sun	1.50×10^{11} m
Nearest star	4.04×10^{16} m
Galactic center	2.2×10^{20} m
Andromeda galaxy	2.1×10^{22} m
Edge of the observable universe	$\sim 10^{26}$ m

Rh __

Text__

SOME MATHEMATICAL SYMBOLS

$=$	equals	∞	infinity
$\approx$	equals approximately	$\lim$	the limit of
$\neq$	is not equal to	Σ	the sum of
$\equiv$	is identical to, is defined as	$\int$	the integral of
$>$	is greater than	Δx	the change or difference in x
$\gg$	is much greater than	$\lvert x \rvert$	the absolute value or magnitude of x
$\geq$	is greater than or equal to	x_{av}	the average value of x
$<$	is less than	$x!$	x factorial
$\ll$	is much less than	$\ln x$	natural logarithm of x
$\leq$	is less than or equal to	$f(x)$	a function of x
$\sim$	is of the order of magnitude of	df/dx	the derivative of f with respect of x
$\propto$	is proportional to	$\partial f/\partial x$	the partial derivative of f with respect to x

SI PREFIXES

Factor	*Prefix*	*Symbol*	*Factor*	*Prefix*	*Symbol*
10^{24}	yotta	Y	10^{-1}	deci	d
10^{21}	zetta	Z	10^{-2}	centi	c
10^{18}	exa	E	10^{-3}	milli	m
10^{15}	peta	P	10^{-6}	micro	μ
10^{12}	tera	T	10^{-9}	nano	n
10^{9}	giga	G	10^{-12}	pico	p
10^{6}	mega	M	10^{-15}	femto	f
10^{3}	kilo	k	10^{-18}	atto	a
10^{2}	hecto	h	10^{-21}	zepto	z
10^{1}	deka	da	10^{-24}	yocto	y

THE GREEK ALPHABET

Alpha	A	α	Iota	I	ι	Rho	P	ρ
Beta	B	β	Kappa	K	κ	Sigma	Σ	σ
Gamma	Γ	γ	Lambda	Λ	λ	Tau	T	τ
Delta	Δ	δ	Mu	M	μ	Upsilon	Y	υ
Epsilon	E	ϵ	Nu	N	ν	Phi	Φ	ϕ, φ
Zeta	Z	ζ	Xi	Ξ	ξ	Chi	X	χ
Eta	H	η	Omicron	O	o	Psi	Ψ	ψ
Theta	Θ	θ	Pi	Π	π	Omega	Ω	ω

SOME SELECTED TABLES

Table	*Chapter*	*Page*
SI Base Units	1	2
SI Prefixes	1	2
Some Terminal Speeds in Air	4	72
Coefficients of Friction	5	97
Rotational Inertias (figure)	9	185
Review and Comparison of Translational and Rotational Dynamics	10	221
Comparison of Energy-Related Translational and Rotational Quantities	11	244
Variation of g_0 with Altitude	14	304
Some Escape Speeds	14	310
Some Densities	15	333
Viscosities of Selected Fluids	16	361
The Speed of Sound	19	431
Some Intensities and Sound Levels	19	433
The Lorentz Transformation Equations	20	459
The Lorentz Velocity Transformation	20	461
Some Average Coefficients of Linear Expansion	21	486
Some Molecular Speeds at Room Temperature	22	500
Some Thermal Conductivities and *R*-Values	23	518
Heat Capacities of Some Substances	23	524
Some Heats of Transformation	23	525
Molar Heat Capacities of Gases	23	532

SOME UNITS AND ABBREVIATIONS

ampere	A	kelvin	K
atmosphere	atm	light-year	ly
British thermal unit	Btu	liter	L
calorie (physical)	cal	meter	m
calorie (nutritional)	Cal	mile	mi
coulomb	C	minute	min
day	d	mole	mol
degree Celsius	°C	newton	N
degree Fahrenheit	°F	ohm	Ω
electron-volt	eV	pascal	Pa
farad	F	pound	lb
foot	ft	radian	rad
gauss	G	revolution	rev
gram	g	second	s
henry	H	tesla	T
hertz	Hz	unified atomic mass unit	u
horsepower	hp	volt	V
hour	h	watt	W
inch	in.	weber	Wb
joule	J	year	y

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.